

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:40:24 (Search time 27 Seconds)

(without alignments)  
922.640 Million cell updates/sec

Title: US-09-673-400a-38

Sequence: 1 ARAPLDMFRRLRLSADPHA.....YRGFLAEKRVQWEDEDF 144

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processings: Minimum Match 0%

Maximum Match 100%

Database:

SPREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	397	52.0	113	11	Q9D1F0
2	386.5	50.7	100	11	Q9DC23
3	378.5	49.6	112	11	Q9D610
4	150	19.7	704	13	Q98SV9
5	137	18.0	325	4	Q96A68
6	137	18.0	342	4	Q9UPV1
7	126.5	16.6	303	4	Q9H7A0
8	122	16.0	371	13	Q93Z83
9	121	15.9	853	3	Q00833
10	105.5	13.8	1252	3	Q96VA3
11	101	13.2	837	3	Q9HFR8
12	97	12.7	231	11	Q9E011
13	85	11.1	639	3	Q9UVC2
14	84	11.0	406	3	Q9UVD9
15	83	10.9	400	6	Q95KK1
16	79	10.4	246	16	Q98BE8

17	78	10.2	223	2	Q45488
18	77.5	10.2	499	16	Q9A8K5
19	74.5	9.8	464	10	Q48967
20	73.5	9.6	589	16	Q9KE11
21	73	9.6	853	1	Q30772
22	72	9.4	183	2	Q93562
23	71.5	9.4	485	2	Q9XAF0
24	71.5	9.4	550	5	Q19680
25	70.5	9.2	180	16	Q9PAL1
26	70.5	9.2	202	4	Q9NVZ0
27	70.5	9.2	202	4	Q9NVZ2
28	70.5	9.2	443	3	Q06098
29	70.5	9.2	688	2	Q9FBZ8
30	70.5	9.2	797	5	Q9NKN3
31	70	9.2	1442	12	Q42066
32	69.5	9.1	476	16	Q9KDN8
33	69.5	9.1	1609	10	Q04648
34	69	9.0	763	2	Q9KR25
35	69	9.0	2506	12	Q9WJD2
36	68.5	9.0	348	2	Q9S213
37	68.5	9.0	471	17	Q9HMA9
38	68.5	9.0	558	2	Q9WXR0
39	68.5	9.0	1538	10	Q94H26
40	68	8.9	215	2	Q93U96
41	68	8.9	250	16	Q910F9
42	68	8.9	437	2	Q9L1X2
43	68	8.9	467	5	Q967X8
44	68	8.9	940	11	Q35157
45	68	8.9	1113	2	Q9L249

## ALIGNMENTS

RESULT 1	Q9D1F0	PRELIMINARY;	PRT;	113 AA.
AC	Q9D1F0			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	1110012005RIK PROTEIN (RIKEN CDNA 1110012005 GENE).			
GN	1110012005RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Airakawa T., Hara A., Fukunishi Y., Kono H., Aichi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzaresi J., Sakamoto N.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	*Functional annotation of a full-length mouse CDNA collection.*			
RL	Nature 409:685-690(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

	RESULT 3
O90610	PRELIMINARY; PRT: 112 AA.
ID O9D610	
AC Q9D610;	
DT 01-JUN-2001 (TRMBLrel. 17, Created)	
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)	
DT 01-JUN-2001 (TRMBLrel. 17, last annotation update)	
DE 2900027G03RIK PROTEIN.	
CN 2900027G03RIK.	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX NCBI_Taxid=10090;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;	
RX MEDLINE=21085660; PubMed=11217851;	
RA Kawai J., Shingagawa A., Shibata R., Yoshino M., Itoh M., Ishii Y.,	
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,	
RA Atzawa K., Izawa Y., Nishi K., Kiyosawa H., Kondou S., Yamataka I.,	
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasuka T., Satto R.,	
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,	
RA Fleischmann W., Gaasterland T., Gissi C., Kling B., Kochiya H.,	
RA Kleih P., Lewis S., Matsuo Y., Nikolaev I., Persole G., Quackenbush J.,	
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,	
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,	
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,	
RA Brownstein N.J., Butt C., Fletcher N., Humé D.A., Kamlya M., Lee N.H.,	
RA Gustincich S., Hill D., Hofman M., Hume D.A., Kamiya M., Lee N.H.,	
RA Lyons P., Marchionni L., Mashima J., Mazzarello J., Mombearts P.,	
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,	
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,	
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,	
RA Wysshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohlsuki S.,	
RA Hayashizaki Y.;	
"Functional annotation of a full-length mouse cDNA collection."	
RL Nature 409:685-690(2001).	
EML: AK013603; BAB28923.1; "	
DR MGI:1920115; 2900027G03RIK.	
SQ SEQUENCE 112 AA: 13578 MW: 14809ACB3D50516 CRC64:	
Query Match 49.6%; Score 378.5; DB 11; Length 112;	
Best Local Similarity 64.0%; Pred. No. 2.le-34;	
Matches 73; Conservative 14; Mismatches 24; Indels 3; Gaps 2	
OY 32 MDGRVQLMKALLAGF-LRRARRRMRPRPEPFSDDBDLRPEITVQTGSYMFDNTFS 90	
Db 1 MKRIKLIALMAWELI --II::IIIIIII IIII:::IIIIIII II II II II	
OY 91 NDALVFTELTLTGPALOWIVPIRKESPLNDYDGFLAEMKRVGWEDEDF 144	
Db 59 TDKKRVMLTIRLKRGALOMARVIQTDSPLMNTNSGFLENKEKGEWEDDEF 112	
RESULT 4	
O98SV9 PRELIMINARY; PRT: 704 AA.	
ID O98SV9	
AC O98SV9;	
DT 01-JUN-2001 (TRMBLrel. 17, Created)	
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)	
DT 01-DEC-2001 (TRMBLrel. 19, Last annotation update)	
DE GAG-PROLEASE.	
OS Fuugu rubripes (Japanese pufferfish) (Takifugu rubripes).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Actinopterygii; Neopleterygia; Teleostei; Euteleostei; Neoteleostei;	
OC Acanthomorphae; Acantopterygii; Percormorpha; Tetraodontiformes;	
OC Tetraodonitidae; Takifugu.	
OX NCBI_TaxID=31033;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TRANSPORT-SUSH1--SAN RETROTRANSPOSON;	
RX MEDLINE=21321259; PubMed=11428463;	

RA Butler M., Goodwin T., Simpson M., Singh M., Poulter R.;  
 RT "Vertebrate LTR retrotransposons of the Tf1/sushi group."  
 RL J. Mol. Evol. 52:260-274(2001).  
 DR EMBL: AF316578; AAG60684.1; -;  
 DR MEROPS; A02.0PM; -;  
 DR InterPro: IPR001969; Asp.-protease.  
 DR InterPro: IPR001995; Asp.-prot.-retrov.  
 DR InterPro: IPR000953; Chromo.  
 DR InterPro: IPR001878; ZnF\_CCHC.  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR SMART: SM00343; ZnF\_C2HC; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE: PS50175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE: PS50013; CHROMO\_2; 1.  
 DR PROTEASE; Zinc-finger.  
 SQ SEQUENCE 704 AA; 76619 MW; 81E2ELC9AB0153B3 CRC64;

Query Match 19.7%; Score 150; DB 13; Length 704;  
 Best Local Similarity 27.7%; Pred. No. 5.5e-08;  
 Matches 43; Conservative 24; Mismatches 52; Indels 36; Gaps 5;

QY 16 ADPHATONSAREGTMDGR-VOLM-----KAL 42  
 DB 4 AODAVRR-TLEARGRLQHNOLLIDINASLOSINTSVTDLSIGMAEQVSPRAEAL 62  
 QY 43 LAGPLPARRMRNPPIPEPTFGDTDRLEPTIVQCSYWF-VDENTFSDALKVFLIT 101  
 DB 63 EVYTVAAAPMEPHPIPERYSGEAGVCASFLLQ-CSLVEFDIQLPLYPDGRKIAFVNV 121  
 QY 102 RLGPALQWVPIRKESPLINDYRGFLMKRVF 136  
 DB 122 LLSGRAQWATAVLENQTPASSFSPEFTALKRVF 156

## RESULT 5

Q96A68 PRELIMINARY; PRT; 325 AA.

AC Q96A68;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ME3 LIKE 1 (PREFERENTIALLY EXPRESSED GENE 10 ORF1).  
 GN ME3L1 OR PEG10.  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;

RA Nakamura Y., Furukawa Y.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Satoh S., Furukawa Y.;  
 RT "Isolation of ME3 like gene 1."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-21218929; PubMed-11318613;  
 RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,  
 RA Kaneko-Ishino T., Ishino F.;  
 RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene  
 RT Located on Human Chromosome 7q21."  
 RL Genomics 73:232-237(2001).  
 DR EMBL: AB049150; BAB68387.1; -;  
 DR EMBL: AB049834; BAB43951.1; -;  
 DE CDNA: FLJ21125 FIS, CLONE CAS06077 (HYPOTHEICAL 32.5 KDA  
 SQ SEQUENCE 325 AA; 36965 MW; 118E4CAF97E2A76 CRC64;

Query Match 18.0%; Score 137; DB 4; Length 325;  
 Best Local Similarity 29.4%; Pred. No. 5.8e-07;

Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

QY 11 RRLSADPHATONSAREGTMDGRVOLMKALLAGPLPARRMRNPPIPEPTFGDTDR 70  
 DB 45 REQVEPTPE-DEDDIELRGA-----AAAAAPPIEEEC--PEDLPKFDGNPDM 92  
 QY 71 LPEFIVQCSYMFVDENTFSDALKVFLITRLGPPALQWVPIRKESPLINDYRGFLA 130  
 DB 93 LAPPMAOCQIFMEKSTDFSVDRVRCFVSMATGRARASAKLENSHILMHYPAFMM 152  
 QY 131 EMKRVF 136  
 DB 153 EMKHVF 158

## RESULT 6

Q9UPV1 PRELIMINARY; PRT; 342 AA.

AC Q9UPV1;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE KIAA1051 PROTEIN (FRAGMENT).  
 GN KIAA1051.  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-99397452; PubMed-10470851;  
 RA Kikuno R., Nagase T., Ishikawa K., Hikosawa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 6:197-205(1999).  
 DR EMBL: AB028974; BA83003.1; -;  
 DR InterPro: IPR001878; ZnF\_CCHC.  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR SMART: SM00343; ZnF\_C2HC; 1.  
 KW Zinc-finger.  
 FT NON\_TER 1  
 SQ SEQUENCE 342 AA; 38769 MW; DC14C265B92541EB CRC64;

Query Match 18.0%; Score 137; DB 4; Length 342;  
 Best Local Similarity 29.4%; Pred. No. 6.2e-07;  
 Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

QY 11 RRLSADPHATONSAREGTMDGRVOLMKALLAGPLPARRMRNPPIPEPTFGDTDR 70  
 DB 62 REQVEPTPE-DEDDIELRGA-----AAAAAPPIEEEC--PEDLPKFDGNPDM 109  
 QY 71 LPEFIVQCSYMFVDENTFSDALKVFLITRLGPPALQWVPIRKESPLINDYRGFLA 130  
 DB 110 LAPPMAOCQIFMEKSTDFSVDRVRCFVSMATGRARASAKLENSHILMHYPAFMM 169  
 QY 131 EMKRVF 136  
 DB 170 EMKHVF 175

## RESULT 7

Q9H7A0 PRELIMINARY; PRT; 303 AA.

AC Q9H7A0;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CDNA: FLJ21125 FIS, CLONE CAS06077 (HYPOTHEICAL 32.5 KDA  
 DE PROTEIN).  
 OS Homo sapiens (Human).

file

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Matnabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohashi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isono T., Sugano S.;  
RT "NEO human cDNA sequencing project."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE-UTERUS; LEIOMYOSARCOMA;  
RA Strausberg R.;  
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK024778; BAB14998.1;  
DR EMBL; BC011679; AAH1679.1;  
KW Hypothetical protein.  
SQ SEQUENCE 303 AA; 32478 MW; 3882DD1459808328 CRC64;

Query Match 16.6%; Score 126.5; DB 4; Length 303;  
Best Local Similarity 32.8%; Pred. No. 7.8e-06;  
Matches 38; Conservative 14; Mismatches 51; Indels 13; Gaps 3;

QY 28 ARGTDGGRVQLKALLAGPL--RPARRRM-----NP1PPEPDGDTDLPEFIVQC 79  
DB 7 ASGTGCG-----KPAERGGLAGHPSSRRHYDFCVPSDGTGDSWLLDRFLAQG 61  
QY 80 SYMFVDENTSDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLAEKRV 135  
DB 62 DVMSEFHEHYQDNISVNCVCEILRLGRQAQWAPYLDGDLPLPDYDELFCQDLKEY 117

## RESULT 8

093283 PRELIMINARY; PRT; 371 AA.

AC 093283;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE GAG POLYPROTEIN.

GN GAG.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthopterygii; Acanthopterygii; Perciformes; Tetraodontiformes;  
OC Tetraodontidae; Takifugu.

OC NCBI\_TaxID=31033;

OX (1)  
RP SEQUENCE FROM N.A.

RC TRANSPOSON-SUSHI-1CH1 RETROTRANSPOSON;  
RA MEDLINE-98382517; PubMed-9714821;

RA Poulter R., Butler M.;  
RT "A retrotransposon family from the pufferfish (fugu) Fugu rubripes.";  
RL Gene 215:241-249(1998).

RN (2)  
RP SEQUENCE FROM N.A.

RC TRANSPOSON-SUSHI-1CH1 RETROTRANSPOSON;  
RA Poulter R.T.M., Butler M.I.;

RL Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF030881; AAC33525.1;  
DR InterPro: IPR001878; Znf.CCHC.

DR Pfam: PF00098; Zf.CCHC; 1.

DR PRINTS; PR00939; C2HCZNFINGER.

DR SMART; SM00343; Znf.C2HC; 1.

DR PolyProtein; Zinc-finger.

SQ SEQUENCE 371 AA; 40540 MW; 617926FC010730AC CRC64;

Query Match 16.0%; Score 122; DB 13; Length 371;  
Best Local Similarity 24.7%; Pred. No. 3.2e-05;  
Matches 44; Conservative 21; Mismatches 67; Indels 46; Gaps 5;

2 RAPTIDMRRRLSADPHATORNSAEARGT-----MDGR 35

DB 14 RPTLPSPLRRVEA--HSAQLSSLOSELTKAFPTTIGELSELQSSQRTSSLSLSNQ 71  
QY 36 VOLKALLAG-----PLRPARRRNP1PPEPDGDTDLPEFIVQC 79  
DB 72 MSAMATVLSIIQKLSDDGGAAPSESLPLSRAPRNPLASPRVFGDDPLGKFLHC-C 130  
QY 80 SYMFVDE-NTSFNDALKVFLITRLTGPALQWVPIRKESPLNDYRGFLAEKRV 136  
DB 131 ELLFRHQRSEVSDKAGVFITSLADKALSWAIAVDDPLRSDYSAPRRERKAVF 188

## RESULT 9

000833 PRELIMINARY; PRT; 853 AA.

AC 000833;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE GAG POLYPROTEIN.

GN GAG.

OS Fusarium oxysporum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; mitosporic Hypocreales; Fusarium.

OC NCBI\_TaxID=5507;

OX (1)  
RP SEQUENCE FROM N.A.

RC STRAIN-F. SP. LYCOPERSICI 42-87;  
RX MEDLINE-96132549; PubMed-8544829;

RA Anaya N., Roncero M.I.;

RT "SkipPy, a retrotransposon from the fungal plant pathogen Fusarium  
oxysporum";

RL Mol. Gen. Genet. 249:637-647(1995).

DR EMBL; L34658; AAA8790.1;  
DR InterPro: IPR001878; Znf.CCHC.

DR Pfam: PF00098; Zf.CCHC; 1.

DR PRINTS; PR00939; C2HCZNFINGER.

DR SMART; SM00343; Znf.C2HC; 1.

DR PolyProtein; Zinc-finger.

SQ SEQUENCE 853 AA; 99422 MW; 1AEDD18CFBBA5B06 CRC64;

Query Match 15.9%; Score 121; DB 3; Length 853;  
Best Local Similarity 34.5%; Pred. No. 0.00012;  
Matches 29; Conservative 12; Mismatches 31; Indels 12; Gaps 2;

QY 47 LRPARRRNP1PPEPDGDTDLPEFIVQCSMFVDENTSFNDALKVFLIT 101  
DB 100 LSAANNGRDPGEVLKPSPEYFDGTPSKLPTLQSRATFYYPQFRDSAKVMYMG 159

QY 102 RLTPALQWVPIRKESPLNDY 125  
DB 160 RLKTRAAQNF-----QPIINDY 176

## RESULT 10

096V43 PRELIMINARY; PRT; 1252 AA.

AC 096V43;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE POLYPROTEIN (FRAGMENT).

OS Aspergillus flavus.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OC NCBI\_TaxID=5059;

OX (1)  
RP SEQUENCE FROM N.A.

RC STRAIN-NRRL 19997; TRANSPOSON-AFRTL-1;  
RX MEDLINE-95314240; PubMed-7793909;

RA McAlpin C.E., Marnett B.;

RT "Construction and characterization of a DNA probe for distinguishing  
strains of Aspergillus flavus";

```

RL Appl. Environ. Microbiol. 61:1068-1072(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL 19997: TRANSPOSON-AFRTL-1.
RA Okubara P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT "AIRT-1, a retrotransposon-like element in the aflatoxin-producing
RT fungus Aspergillus flavus."
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF362957; AAL26311.1; -
FT NON_TER 1252 1252
SQ SEQUENCE 1252 AA; 144752 MW; DEFC5DD8E6C4A8 CRC64;

Query Match 13.8%; Score 105.5; DB 3; Length 1252;
Best Local Similarity 25.4%; Pred. No. 0.01;
Matches 36; Conservative 15; Mismatches 56; Indels 35; Gaps 3;

OY 11 RRLSADPAPATGNS---AAGTMDGROQLKALLAGPLRPAARWRNIPPEPFGD 67
DB 66 RLRNDOOHIAQIDAOVGASAPKDAIGKVKLPKA-----EPPDGT 105
OY 68 TURLPEFIVQCSYMFVDENTFSNDALKYTFILTRTLPALOWVPIYIRK----- 117
DB 106 RSKIDAFELQMMKHIANKNLIDEADKVFISTHLRGAAWMPFPIREYEVVDNWS 165
OY 118 --ESPILNDYRGFLAEKRVFG 137
DB 166 NTRRELFTDSGNLRKHLERFEG 187

RESULT 11
O9HFY8 PRELIMINARY; PRT; 837 AA.
AC O9HFY8:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN.
GN GAG.
OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella
OS cingulata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes Incertae sedis; Phyllostachyales; Phyllostachyaceae;
OC Glomerella.
OX NCBI_TaxID=5457;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-RETROTRANSPOSON CGRET;
RA Zhu P., Oudemans P.V.;
RT "A long terminal repeat retrotransposon Cgret from the phytopathogenic
RT fungus Colletotrichum gloeosporioides on cranberry."
RL Curr. Genet. 0:0-0(2000).
DR EMBL: AF264028; AAG24791.1; -
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.
SQ SEQUENCE 837 AA; 97738 MW; EFID4BC70FD55003 CRC64;

Query Match 13.2%; Score 101; DB 3; Length 837;
Best Local Similarity 29.5%; Pred. No. 0.019;
Matches 36; Conservative 15; Mismatches 59; Indels 12; Gaps 3;

OY 25 SAARAGMDGROVQMLKALL--AGPLRPAAR-----RWRNPIPEPFGDGTDRLP 74
DB 66 NAAALKEIGRARAEQGLKEVATLRAANVISAPEGRELKATNPATFPGTQGLKGH 125
OY 75 IVQCSYMFVDENTFSNDALKYTFILTRTLPALOWVPIYIRKESPLNDYRGFLAEKR 134
DB 126 LVQIRYQAFHMGTFQDTEFVHAATFLRGALAWFEPPL--QOEWLNDPYEKYSOEVRN 183
OY 135 VF 136
DB 184 IF 185

```

```

RESULT 12
O9EQ11 PRELIMINARY; PRT; 231 AA.
AC O9EQ11:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6;
RX MEDLINE-21105984; PubMed-11158386;
RA Volle J.-N., Koerting C., Scharl M.;
RT "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they
RT evolve into new cellular functions?";
RL Mol. Biol. Evol. 18:266-270(2001).
DR EMBL: AF302691; AAG39979.1; -
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 231 AA; 26171 MW; DEAB2A2E624F3974 CRC64;

Query Match 12.7%; Score 97; DB 1; Length 231;
Best Local Similarity 30.9%; Pred. No. 0.011;
Matches 21; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

OY 69 DRLEFIVQCSYMFVDENTFSNDALKYTFILTRTLPALOWVPIYIRKESPLNDYRGF 128
DB 1 DMLGPFMYQCQLFMEKSTROFSVDRIYCVTSLIRAAWATAKLRCTYLMHNNTAF 60
OY 129 LAEMKRVF 136
DB 61 MMELEKHYF 68

RESULT 13
O9QVC2 PRELIMINARY; PRT; 639 AA.
AC O9QVC2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPEPTIDE (PUTATIVE GAG HOMOLOGUE).
GN GAG.
OS Cladosporium fulvum (Fulvia fulva).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetochytrionycetes Incertae Mycosphaerellaceae;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=5499;
RN [1]
RP SEQUENCE FROM N.A.
RC Simpson M.L., Butler M., Poulter R.T.M.;
RT "Functions of the integrase of retrotransposons: the integrase of the
RT Cfr-1 element from Cladosporium fulvum."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-CFT-1 LTR-RETROTRANSPOSON;
RX MEDLINE=92318885;
RA Mchale M.T., Roberts I.N., Noble S.M., Beaumont C., Whitehead M.P.,
RA Seth D., Oliver R.P.;
RT "Cfr-1, an LTR-retrotransposon in Cladosporium fulvum, a fungal
RT pathogen of tomato."
RL Mol. Gen. Genet. 233:337-347(1992).
DR EMBL: AF051915; AAF21677.1; -

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      October 11, 2002, 02:39:14 ; Search time 31 Seconds
              (without alignments)
              515.956 Million cell updates/sec
```

Title:	US-09-673-400A-38
Perfect score:	763
Sequence:	1 ARAPITDMRFRRLSADPHA.....YRGFLAEMKRVFGMEDEDF 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database : A\_Geneseq.032802.\*

1:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	763	100.0	144	20	AAV5927	Human myometrium t
2	566	74.2	113	22	AAB60475	Human cell cycle a
3	549.5	72.0	142	22	ABG12205	Novel human diagno
4	350	45.9	106	21	AAG04029	Human secreted pro
5	335	43.9	126	20	AAV25512	Human mature alpha
6	335	43.9	146	20	AAV26051	Human alpha helica
7	335	43.9	146	22	AAB60492	Human cell cycle a
8	328	43.0	66	21	AAG02068	Human secreted pro
9	316	41.4	110	20	AAV26054	Human zalphal epit
10	310.5	40.7	129	22	ABG12203	Novel human diagno
11	297	38.9	80	20	AAV26056	Human Zalphal epit

12	285	27.4	67	21	AAG02067	Human secreted pro
13	227	29.8	42	22	ABG12204	Novel human diagno
14	201	26.3	65	20	AAV26058	Human zalphal epit
15	197	25.8	184	21	AAV94673	Human zsi983 prote
16	194	25.4	168	21	AAV94674	Human zsi983 matur
17	183	24.0	74	20	AAV26053	Human zalphal epit
18	182	23.9	51	20	AAV26057	Human zalphal epit
19	164	21.5	44	20	AAV26055	Human zalphal epit
20	157	20.6	79	21	AAV94650	Human zsi983 epitc
21	157	20.6	98	21	AAV94651	Human zsi983 epitc
22	154	20.2	65	21	AAV94689	Human zsi983 epitc
23	139.5	18.3	1607	22	ABG15059	Novel human diagno
24	137	18.0	342	21	AAAB2148	Human ORFX ORF1912
25	134	17.6	57	21	AAV94693	Human zsi983 epitc
26	134	17.6	76	21	AAV94694	Human zsi983 epitc
27	131	17.2	43	21	AAV94652	Human zsi983 epitc
28	85	11.1	45	20	AAV26052	Human zalphal epit
29	76.5	10.0	676	22	ABG15498	Novel human diagno
30	76	10.0	41	21	AAV94688	Human zsi983 epitc
31	71.5	9.4	397	22	AAG91310	C glutamicum prote
32	70.5	9.2	202	22	ABBS0194	Human transcriptio
33	70.5	9.2	202	22	ABAB2808	Human protein sequ
34	70.5	9.2	202	22	ABAB2988	Human protein sequ
35	70.5	9.2	243	21	AAAB3610	Human cancer assoc
36	70.5	9.2	245	22	AAV75627	Human colon cancer
37	70.5	9.2	688	22	AAAB07559	Protein encoded by
38	70.5	9.2	2386	22	ABG04328	Novel human diagno
39	70	9.2	358	22	AAAB79919	Corynebacterium gl
40	70	9.2	358	22	AAAB80051	Corynebacterium gl
41	70	9.2	363	22	AAAG92156	C glutamicum prote
42	69	9.0	77	21	AAAB32065	Human secreted pro
43	69	9.0	77	21	AAAB32066	Human secreted pro
44	69	9.0	272	21	AAAG5807	Arabidopsis thailia
45	69	9.0	348	22	AAU28026	Novel human secret

## ALIGNMENTS

XX	RESULT 1
XX	AA559927
XX	ID AA559927 standard; protein; 144 AA.
XX	AC AA559927;
XX	DT 28-JAN-2000 (first entry)
XX	DE Human myometrium tumour EST encoded protein 7.
XX	KW Myometrium; tumour; human; expressed sequence tag; EST; uterine myoma;
XX	KW treatment; carcinoma; cancer; gene therapy.
XX	OS Homo sapiens.
XX	DE19817947-A1.
XX	PD 28-OCT-1999.
XX	PE 17-APR-1998; 98DE-1017947.
XX	PR 17-APR-1998; 98DE-1017947.
XX	PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX	PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX	DR WPI: 1999-602380/52.
XX	DR N-PSDB; AA241965.
XX	PT New nucleic acid sequences expressed in uterine myoma, and derived
XX	PT polypeptides, for treatment of uterine carcinoma and identification of
XX	PT therapeutic agents





XX	Dmanac RT, Liu C, Tang YT:
PI	
XX	DR
XX	WPI; 2001-639362/73.
XX	N-PSDB; AAS76392.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
XX	ClaIm 20; SEQ ID No 42564; 103pp: English.
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (I) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	spectification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 142 AA;
Query Match	72.0%: Score 549.5; DB 22; Length 142;
Best Local Similarity	74.8%; Pred. No. 5.8e-58;
Matches 107; Conservative 12; Mismatches 19; Indels 5; Gaps 2	
OY	6 LDMRRRLSLADPHNMQNRNSAEARGMNGR----VOLMKALLAGPLRPARRMRNPPIP 61
Db	1 ISLRILDLGGACTHSAATEASDR-SEDLKMTVCALDIALALPRLPRTRRRRNPIPP 59
OY	62 ETFDDDTLRLPEFIYOTCSYMFVDENTSGNDALKVTFLITRLTGAPALQWVIPYRKESPL 121
Db	60 ETFDDDTLRLPEFIYOTCSYMYVDENTFSSDALKVTFLITRLTGAPALQWVIPYRKESPL 119
OY	122 LNDYRGFLAEMKRVFGEWEDEF 144
Db	120 LNDYRGFLAEMKRVFGEWEDEF 142
RESULT 4	
AAG04029	
ID	AAG04029 standard; Protein; 106 AA.
AC	AAG04029;
DT	06-OCT-2000 (first entry)
DE	Human secreted protein, SEQ ID NO: 8110.
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KX	gene therapy; chromosome mapping.
OS	Homo sapiens.
PX	EP1033401-A2.
PD	06-SEP-2000.

PF	21-FEB-2000;	2000EP-0200610.	
XX			
PR	26-FEB-1999;	99US-0122487.	
PA	(GEST )	GENSET.	
PI	Dumas Milne Edwards J,	Duclert A, Giordano J;	
XX			
PI	WPI: 2000-500381/45.		
DR	N-PSDB; AAC04035.		
XX			
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for		
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for		
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -		
XX			
PS	Claim 13; SEQ ID 8110; 71bp + CD-ROM; English.		
XX			
CC	The present sequence is a polypeptide encoded by one of a large number		
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs		
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30		
CC	different tissues. EST sequences usually correspond mainly to the 3'		
CC	untranslated region (UTR) of the mRNA because they are often obtained		
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for		
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in		
CC	some cases where longer cDNA sequences have been obtained, the full 5'		
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'		
CC	ends and can therefore be used to obtain full length cDNAs and genomic		
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and		
CC	chromosome mapping procedures. They are used to obtain upstream		
CC	regulatory sequences and to design expression and secretion vectors.		
XX			
SO	Sequence	106 AA;	
	Query Match	45.9%;	Score 350;
	Best Local Similarity	89.3%;	Pred. No. 3.8e-34;
	Matches 67;	Conservative 3;	Mismatches 5;
			Indels 0;
			Gaps 0;
OY	32 MDGRVQLMKALLAGLPAPRRMRNPDPPEFDGDTDLPEFIVOTCSYMFVDENTFSN 91		
DB	1 MDGRVQLIKALALDIRPTRMRNPDPPEFDGDTDLPEFIVOTCSYMFVDENTFSN 60		
OY	92 DALKVTFLITRLTGP 106		
DB	61 DALKVTFLITRLTGP 75		
	RESULT 5		
	AAV25512		
ID	AAV25512 standard; Protein; 126 AA.		
XX			
XX	AAV25512;		
XX			
DT	28-SEP-1999 (first entry)		
XX			
DDE	Human mature alpha helical protein-1 zalpal.		
XX			
KW	Human alpha helical protein-1; Zalpal; helical cytokine; skin tone;		
KW	growth hormone; erythropoietin; leptin; interleukin-10; gene therapy;		
KW	chromosome Xq77.3; FMR1; Fragile-X syndrome; cosmetic improvement;		
KW	hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;		
KW	abnormal proliferative disorder; cancer; connective tissue dysfunction;		
KW	epidermal system; elasticity.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Protein	1..127	
FT		/label= zalpal	
FT		/note= "mature protein fragment, no start codon given"	
XX			
PN	W09929720-A2.		
XX			
PD	17-JUN-1999.		

```

XX PF 10-DEC-1998; 98WO-US26273.
XX PR 10-DEC-1997; 97US-0987926.
XX (ZYMO ) ZYMOGENETICS INC.
XX PA Conklin DC, Lok S, Parrish J;
XX PI WPI; 1999-385572/32.
XX DR N-PSDB; AAX80687.
XX PT Mammalian alpha helical protein-1, designated Zalphal
XX PS Claim 5; Page 69; 73pp; English.
XX PS
XX CC The present sequence represents the mature protein fragment of a novel
XX CC alpha helical protein-1 designated Zalphal from human pituitary gland
XX CC cDNA library. Zalphal is predicted to be a four-helical protein similar
XX CC to the family of helical cytokines represented by growth hormone,
XX CC erythropoietin, leptin and interleukin-10. Zalphal gene was mapped to
XX CC chromosome Xq27.3, in close proximity to FMRI, a gene linked to fragile-X
XX CC syndrome. Its transcription levels were found to be reduced or absent in
XX CC fragile-X patients. Zalphal transcripts were found at high levels in
XX CC pituitary and aorta, and lower levels in brain, kidney, pancreas,
XX CC prostate, etc. Zalphal is supposed to be responsible for the
XX CC hypothalamic-pituitary-gonadal axis dysfunction and connective tissue
XX CC dysfunctions in fragile-X patients. The Zalphal DNA is used in gene
XX CC therapy for treating patients having a mutated Zalphal gene or lacking
XX CC the gene. Probes derived from Zalphal gene can be used to check
XX CC abnormalities on X chromosome. Zalphal protein may be useful in the
XX CC treatment of fragile-X syndrome and abnormal proliferative disorders e.g.
XX CC cancer. It can also be used for the growth, differentiation, maintenance
XX CC and survival of connective tissues, particularly cardiovascular and
XX CC epidermal systems and in imparting cosmetic improvements to normal
XX CC connective tissues such as enhancement of skin tone and elasticity.
XX SQ Sequence 126 AA:
XX
XX Query Match 43.9%; Score 335; DB 20; Length 126;
XX Best Local Similarity 59.6%; Pred. No. 3e-32;
XX Matches 62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;
XX
XX QY 40 KALLAGLPAPARRMRNDIPPEFTDGTDLRPEFIVQTCSTMFVDETFSSNDALKVTL 99
XX :| | :| | :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX 17 RASLRQVRPPS---CPVPEPEFNGESSRLPEFIVQTASYMLVNERFCNDAMKAVFL 72
XX
XX QY 100 ITRLTGPALOWIVIPYIKRESPLNDYRGFLAEMKRVFGWEDED 143
XX :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 73 ISLTGEAEWVVPYIEMDSPILGDYRAFLDEMKOCFGWDDDED 116
XX
XX Db
XX
XX RESULT 6
XX AAY26051 standard; Protein: 146 AA.
XX
XX AC AAY26051;
XX
XX XX 28-SEP-1999 (first entry)
XX
XX DE Human alpha helical precursor protein-1, Zalphal.
XX
XX KW Human alpha helical precursor protein-1; Zalphal; helical cytokine;
XX KW growth hormone; erythropoietin; leptin; chromosome Xq27.3; FMRI;
XX KW fragile-X syndrome; interleukin-10; connective tissue dysfunction;
XX KW abnormal proliferative disorder; cancer; epidermal system; skin tone;
XX KW hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;
XX KW cosmetic improvement; elasticity.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..20

```

```

FT FT /label= signal_peptide
FT FT 21..146
FT FT /label= Zalphal
FT FT 23..37
FT FT /label= Helix_A
FT FT 53..67
FT FT /label= Helix_B
FT FT 82..96
FT FT /label= Helix_C
FT FT 118..132
FT FT /label= Helix_D
XX
XX PN WO929720-A2.
XX
XX PD 17-JUN-1999.
XX
XX PF 10-DEC-1998; 98WO-US26273.
XX PR 10-DEC-1997; 97US-0987926.
XX (ZYMO ) ZYMOGENETICS INC.
XX PA Conklin DC, Lok S, Parrish J;
XX PI WPI; 1999-385572/32.
XX DR N-PSDB; AAX80687.
XX PT Mammalian alpha helical protein-1, designated Zalphal
XX PS Example 2; Pages 60-61; 73pp; English.
XX
XX CC The present sequence represents an alpha helical precursor protein-1
XX CC designated Zalphal from human pituitary gland cDNA library. Zalphal
XX CC is predicted to be a four-helical protein similar to the family of
XX CC helical cytokines represented by growth hormone, erythropoietin, leptin
XX CC and interleukin-10. Zalphal gene was mapped to chromosome Xq27.3, in
XX CC close proximity to FMRI, a gene linked to fragile-X syndrome. Its
XX CC transcription levels were found to be reduced or absent in fragile-X
XX CC patients. Zalphal transcripts were found at high levels in pituitary and
XX CC aorta, and lower levels in brain, kidney, pancreas, prostate, etc.
XX CC It is supposed to be responsible for the hypothalamic-pituitary-
XX CC gonadal axis dysfunction and connective tissue dysfunctions in fragile-X
XX CC patients. Zalphal protein may be useful in the treatment of fragile-X
XX CC syndrome and abnormal proliferative disorders e.g. cancer. It can also be
XX CC used for the growth, differentiation, maintenance and survival of
XX CC connective tissues, particularly cardiovascular and epidermal systems
XX CC and in imparting cosmetic improvements to normal connective tissues such
XX CC as enhancement of skin tone and elasticity.
XX
XX SQ Sequence 146 AA:
XX
XX Query Match 43.9%; Score 335; DB 20; Length 146;
XX Best Local Similarity 59.6%; Pred. No. 3.6e-32;
XX Matches 62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;
XX
XX QY 40 KALLAGLPAPARRMRNDIPPEFTDGTDLRPEFIVQTCSTMFVDETFSSNDALKVTL 99
XX :| | :| | :| | | | | | | | | | | | | | | | | | | | | | | | | |
XX 37 RASLRQVRPPS---CPVPEPEFNGESSRLPEFIVQTASYMLVNERFCNDAMKAVFL 92
XX
XX QY 100 ITRLTGPALOWIVIPYIKRESPLNDYRGFLAEMKRVFGWEDED 143
XX :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 93 ISLTGEAEWVVPYIEMDSPILGDYRAFLDEMKOCFGWDDDED 136
XX
XX Db
XX
XX RESULT 7
XX AAB60492 standard; Protein: 146 AA.
XX
XX AC AAB60492;
XX
XX XX 24-APR-2001 (first entry)
XX
XX DE Human cell cycle and proliferation protein CCYR-40, SEQ ID NO:40.

```

XX	Cell cycle and proliferation protein; CCYPR; human; agonist;
KM	antagonist; gene therapy; detection; gene therapy;
KW	transgenic animal disease model; immune disorder;
KW	developmental disorder; cell signalling disorder;
KW	cell proliferative disorder; cancer; tumour; anaemia;
KW	arteriosclerosis; asthma; allergy; diabetes mellitus;
KW	menstrual cycle disorder; bacterial infection.
XX	
OS	Homo sapiens.
XX	
PN	MO200107471-A2.
XX	
PD	01-FEB-2001.
XX	
XX	21-JUL-2000; 2000MO-US19948.
XX	
PR	21-JUL-1999; 99US-0145075.
PR	08-SEP-1999; 99US-0153129.
PR	10-NOV-1999; 99US-0164647.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
PI	Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI	Azizmezi Y, Yang J, Lu DM, Baughn MR, Patterson C, Shah P;
XX	
DR	WPI: 2001-112727/12.
XX	
DR	N-PSDB; AAF59629.
XX	
PT	Human cell cycle and proliferation proteins and polynucleotides are
PT	used to treat, diagnose and prevent immune, developmental and cell
PT	signalling disorders and cell proliferative disorders. Including cancer -
XX	
PS	Disclosure: Page 150; 205pp; English.
XX	
CC	Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC	proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC	CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC	associated with decreased expression of functional CCYPR, while CCYPR
CC	antagonists are used to treat diseases or conditions associated with
CC	overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC	to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC	radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC	compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC	that specifically bind to CCYPR, and in drug screening methods to
CC	identify compounds that modulate the activity of CCYPR. CCYPR
CC	nucleotides can be used to generate transgenic animal models of human
CC	disease, and can be used in gene therapy in target cells with genetic
CC	abnormalities with respect to the expression of CCYPR for the
CC	treatment or prevention of a disorder associated with CCYPR.
CC	Diseases which can be diagnosed, treated and prevented using CCYPR
CC	proteins, nucleic acids, agonists or antagonists include immune,
CC	developmental and cell signalling disorders, and cell proliferative
CC	disorders including cancer. Specific examples of these disorders
CC	include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC	diabetes mellitus, disorders of the menstrual cycle and infections
CC	caused by bacteria.
XX	
SO	Sequence 146 AA;
XX	
Query Match	43.9%; Score 335; DB 22; Length 146;
Best Local Similarity	59.6%; Pred. No. 3.6e-32;
Matches 62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;	
OY	40 KALLAGPLPARRRRNRNPIPEPTDGTDRLPETIVQCSFMFDENTFSMDALKYTRL 99
DB	37 RASTLRQVRPPS---CVPPEPTENGESSRLPEETIVQASVMTLVNENRFDAMKVAEL 92
OY	100 ITRLGSPALOWIVYIRKESPLNIVRGRLAKMKVFGCEED 143
DB	93 ISLTGEAEWVYPIETENDSPILGDRFAFLDMKQCFGMDDED 136

RESULT	8
AAG02068	
ID	AAG02068 standard; Protein: 66 AA.
XX	
AC	AAG02068:
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein, SEQ ID NO: 6149.
XX	
KM	Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX	
KW	gene therapy; chromosome mapping.
OS	Homo sapiens.
XX	
FN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PE	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST ) GENSET.
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
DR	WPI: 2000-500381/45.
DR	N-PSDB: AAC02074.
XX	
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX	
PS	Claim 13: SEQ ID 6149; 71pp + CD-ROM; English.
XX	
CC	The present sequence is a polypeptide encoded by one of a large number
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30
CC	different tissues. EST sequences usually correspond mainly to the 3'
CC	untranslated region (UTR) of the mRNA because they are often obtained
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC	those cases where longer cDNA sequences have been obtained, the full 5'
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC	ends and can therefore be used to obtain full length cDNAs and genomic
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC	chromosome mapping procedures. They are used to obtain upstream
CC	regulatory sequences and to design expression and secretion vectors.
XX	
SQ	Sequence 66 AA:
Query Match	43.0%; Score 328; DB 21; Length 66;
Best Local Similarity	95.5%; Pred. No. 9.2e-32;
Matches 63; Conservative 0; Mismatches 3; Indels 0; Gaps 0	
OY	32 MDGRVOLMKALLAGPLPARARRMRNDIPPEPFDCDTRLPEFTVTCSYMFVDENTFSN 91   Db 1 MXGRVOLMKALLAXPLRRPARRRMRNDIPPPEFDGDTDLRLPEFTVTSYMFVDENTFSN 60
OY	92 DALKYT 97           Db 61 DALKVT 66
RESULT 9	
AAV26054	
ID	AAV26054 standard; Protein: 110 AA.
XX	
AC	AAV26054:
DT	28-SEP-1999 (first entry)
XX	



XX 17-JUN-1999.  
 PD 10-DEC-1998; 98WO-US26273.  
 XX 10-DEC-1997; 97US-0987926.  
 PR 10-DEC-1997; 97US-0987926.  
 XX (ZYMO ) ZYMOGENETICS INC.  
 PA (ZYMO ) ZYMOGENETICS INC.  
 XX Conklin DC, Lok S, Parrish J;  
 PI WPI; 1999-385572/32.  
 XX Mammalian alpha helical protein-1, designated zalphal  
 PT Claim 10; Page 72; 73pp; English.  
 XX The present sequence is an epitope-bearing protein derived from  
 CC human alpha helical protein-1, zalphal. The protein comprises  
 CC helices B, C and D of mature zalphal. It is used to raise specific  
 CC antibodies which can be used for detection and purification of zalphal.  
 CC The zalphal protein may be useful in the treatment of Fragile-X  
 CC syndrome and abnormal proliferative disorders e.g. cancer. It can also be  
 CC used for the growth, differentiation, maintenance and survival of  
 CC connective tissues, particularly cardiovascular and epidermal systems  
 CC and in imparting cosmetic improvements to normal connective tissues such  
 CC as enhancement of skin tone and elasticity.  
 XX Sequence 80 AA:  
 SQ  
 Query Match 38.9%; Score 297; DB 20; Length 80;  
 Best Local Similarity 66.2%; Pred. No. 6.2e-28;  
 Matches 53; Conservative 12; Mismatches 15; Indels 0; Gaps 0;  
 QY 60 FPFPGDPTDLPEFIVGCSYMFVDETFESNDALKTFLITRLGPALOWIPIRKES 119  
 DB 1 FPFENGESSRLPEFIVGTASYMLVFNENRFDNKAFLISLTGAEVWVPIEDMS 60  
 QY 120 PLNDYRGFLAMKRVFGME 139  
 DB 61 PILGDYRAFLDMKOCFGMD 80

RESULT 12  
 AAG02067  
 ID AAG02067 standard; Protein: 67 AA.  
 AC AAG02067;  
 XX 06-OCT-2000 (first entry)  
 DE Human secreted protein. SEQ ID NO: 6148.  
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX EP1033401-A2.  
 PN 06-SEP-2000.  
 PD 21-FEB-2000; 2000EP-0200610.  
 PF 26-FEB-1999; 99US-0122487.  
 PR (GEST ) GENSET.  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI; 2000-500381/45.  
 DR N-PSDB; AAC02073.  
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 13; SEQ ID 6148; 71pp + CD-ROM; English.  
 XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX Sequence 67 AA:  
 SQ  
 Query Match 37.4%; Score 285; DB 21; Length 67;  
 Best Local Similarity 86.9%; Pred. No. 1.4e-26;  
 Matches 53; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 32 MDGRVQLKALLAGLPAPARRMRNPFPETFDGDTDLPEFIVGCSYMFVDETFESN 91  
 DB 1 MDGRVQLKALLALPAPARRMRNPFPETFDGDTDLPEFIVGCSYMFVDETFESN 60  
 QY 92 D 92  
 DB 61 E 61

RESULT 13  
 ABG12204  
 ID ABG12204 standard; Protein: 42 AA.  
 AC ABG12204;  
 XX 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #12195.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS76391.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 20; SEQ ID NO 42563; 103pp; English.  
 XX

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 42 AA:  
 SQ  
 Query Match 29.8%; Score 227; DB 22; Length 42;  
 Best Local Similarity 97.6%; Pred. No. 7.1e-20;  
 Matches 41; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 103 LTGPALQWVPIPIKESPLNDYRGFLAMKRVGMEDEDE 144  
 Db 1 LTGPALQWVPIPIKESPLNDYRGFLAMKRVGMEDEDE 42  
 RESULT 14  
 AAY26058  
 ID AAY26058 standard; Protein: 65 AA.  
 XX  
 AC AAY26058;  
 XX  
 DT 28-SEP-1999 (first entry)  
 XX  
 DE Human Zalphal epitope-bearing protein fragment 7.  
 XX  
 KW Alpha helical protein-1; Fragile-X syndrome; cardiovascular system;  
 KW connective tissue; abnormal proliferative disorder; cancer; skin tone;  
 KW epidermal system; cosmetic improvement; skin tone; elasticity; Zalphal;  
 KW epitope.  
 XX  
 KW Homo sapiens.  
 OS  
 XX MO9929720-A2.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 10-DEC-1998; 98WO-US26273.  
 XX  
 PR 10-DEC-1997; 97US-0987926.  
 XX  
 PA (ZYMO) ZYMOGENETICS INC.  
 XX  
 PI Conklin DC, Lok S, Parrish J;  
 XX  
 DR WPI; 1999-385572/32.  
 XX  
 XX Mammalian alpha helical protein-1, designated Zalphal  
 XX  
 PS Claim 10; Pages 72-73; 73pp; English.  
 CC  
 CC The present sequence is an epitope-bearing protein derived from  
 CC human alpha helical protein-1, Zalphal. It is used to raise specific  
 CC antibodies which can be used for detection and purification of Zalphal.  
 CC The Zalphal protein may be useful in the treatment of Fragile-X  
 CC syndrome and abnormal proliferative disorders e.g. cancer. It can also be

CC used for the growth, differentiation, maintenance and survival of  
 CC connective tissues, particularly cardiovascular and epidermal systems  
 CC and in imparting cosmetic improvements to normal connective tissues such  
 CC as enhancement of skin tone and elasticity.  
 CC  
 XX Sequence 65 AA:  
 SQ  
 Query Match 26.3%; Score 201; DB 20; Length 65;  
 Best Local Similarity 63.6%; Pred. No. 1.7e-16;  
 Matches 35; Conservative 9; Mismatches 11; Indels 0; Gaps 0;  
 Oy 89 FSNALRYTFLITRLTGPALQWVPIPIKESPLNDYRGFLAMKRVGMEDEDE 143  
 Db 1 FSNALRYTFLITRLTGPALQWVPIPIKESPLNDYRGFLAMKRVGMEDEDE 55  
 RESULT 15  
 AAY94673  
 ID AAY94673 standard; Protein: 184 AA.  
 XX  
 AC AAY94673;  
 XX  
 DT 01-DEC-2000 (first entry)  
 XX  
 DE Human zsig83 protein sequence.  
 XX  
 KW Alpha-helical protein; zsig83; cell growth; differentiation; cancer;  
 KW proliferation; chromosome 22q13.1-q13.2; cytostatic; vulnerey;  
 KW degenerative condition; metastasis; wound healing.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX FH Peptide 1..16  
 XX FT /note= "Putative signal sequence"  
 XX FT 17..184  
 XX FT /label= zsig83  
 XX FT 38..54  
 XX FT /note= "SH3-binding domain"  
 XX FT 57..71  
 XX FT /label= Helix\_A  
 XX FT 79..97  
 XX FT /label= Helix\_B  
 XX FT 76..81  
 XX FT /note= "Hydrophilic region"  
 XX FT 90..95  
 XX FT /note= "Hydrophilic region"  
 XX FT 113..121  
 XX FT /label= Helix\_C  
 XX FT 119..124  
 XX FT /note= "Hydrophilic region"  
 XX FT 126..135  
 XX FT /label= Helix\_D  
 XX FT 144..154  
 XX FT /label= Helix\_E  
 XX FT 142..147  
 XX FT /note= "Hydrophilic region"  
 XX FT 178..183  
 XX FT /note= "Hydrophilic region"  
 XX  
 PN WO200050594-A2.  
 XX  
 XX 31-AUG-2000.  
 XX  
 PD 25-FEB-2000; 2000WO-US04816.  
 XX  
 PF 26-FEB-1999; 99US-0259131.  
 XX  
 PR (ZYMO) ZYMOGENETICS INC.  
 XX  
 PA Presnell SR;  
 XX  
 PI WPI; 2000-572091/53.  
 XX  
 DR

DR N-PSDB: AAA28032, AAA28037.

XX Alpha-helical protein zsi983, its antibodies and the polynucleotide  
PT encoding the protein useful for treating disorders associated with  
PT abnormal cell growth e.g. cancer and agonists useful for treating  
PT wounds

XX Claim 1: Page 73-74; 83pp; English.

CC This invention relates to a novel human alpha-helical protein designated  
CC zsi983. zsi983 plays a role in the process of cell growth.  
CC differentiation, or proliferation. The zsi983 gene is located on  
CC chromosome 22 at position 22q13.1-q13.2. Included in the invention are  
CC polynucleotide sequences encoding the zsi983 protein, expression vectors  
CC containing the zsi983 DNA sequence, a cultured cell containing the  
CC expression vector, and antibodies specific to the zsi983 protein. The  
CC zsi983 protein contains 5 alpha helix regions (represented by sequences  
CC AAY94677-Y94681) and also contains epitope bearing regions (represented  
CC by sequences AAY94688-Y94698) to which the antibodies are directed. The  
CC protein exhibits cytostatic and vulneray activity. The zsi983 protein  
CC and nucleotide sequences and antibodies are used for treating disorders  
CC associated with abnormal cell growth e.g. cancer, degenerative  
CC conditions and metastasis. The zsi983 protein and its agonists or  
CC antagonists are useful for promoting wound healing. The zsi983 DNA  
CC sequence can be used to identify defective zsi983 genes and may therefore  
CC be used as a diagnostic indicator of cancer.  
CC The present sequence represents the human zsi983 protein.

XX Sequence 184 AA;

Query Match 25.8%; Score 197; DB 21; Length 184;  
Best Local Similarity 36.0%; Pred. No. 1.9e-15;

Matches 40; Conservative 25; Mismatches 40; Indels 6; Gaps 2;

QY	29	RGTMGRVQL--WKALLAGPLRPARRRNPFPETFDGTDRLPEFIVQCSYFVD	85
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
DB	15	RARIPGALQITPPISSTISNCTRPMT--TPPTSLPEPFSGDPGRLAGFLMDRFMIFQ	71
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
QY	86	ENFESNDALKVTPLITRTLPALQWVTPYTRKESPLNDYRGFLAEKRVF	136
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
DB	72	ASRPGEAEHVAFLVSRLTGAEKVALPHMOPDSPLRNNTYOGFLAEIARTY	122
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	

Search completed: October 11, 2002, 02:42:03  
Job time : 32 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:40:54 ; Search time 20 Seconds

(without alignments)  
691.843 Million cell updates/sec

Title: US-09-673-400A-38

Perfect score: 763  
Sequence: 1 ARAPTLDMRRRRRLSADPHA.....YRGFLAEKKRVFGWEDEDF 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	15.9	853	2	S60178
2	104	13.6	457	2	T18347
3	85	11.1	639	2	S23569
4	78	10.2	223	2	JCS323
5	77.5	10.2	499	2	E87416
6	77	10.1	232	2	JE0163
7	74.5	9.8	464	2	T01324
8	74	9.7	414	2	F70778
9	73.5	9.6	589	2	G83758
10	73	9.6	853	2	T08162
11	71.5	9.4	485	2	T35628
12	71.5	9.4	550	2	T21194
13	70.5	9.2	180	2	D82550
14	70.5	9.2	443	2	S59771
15	70	9.2	1442	2	T42607
16	69.5	9.1	476	2	E83796
17	69.5	9.1	1609	2	T01797
18	69	9.0	348	1	R5HUL3
19	69	9.0	1487	1	EDBE1
20	68.5	9.0	348	2	T36890
21	68.5	9.0	471	2	B84412
22	68.5	9.0	586	2	AF1136
23	68.5	9.0	1515	1	S51863
24	68	8.9	250	2	A83310
25	68	8.9	1415	2	C86438
26	67.5	8.8	201	2	AB2216
27	67.5	8.8	316	2	S25843
28	67.5	8.8	971	2	S28833
29	67	8.8	729	2	F86308

30	66.5	8.7	269	2	G90261
31	66.5	8.7	325	2	G72020
32	66.5	8.7	337	2	A53041
33	66.5	8.7	430	2	F87112
34	66.5	8.7	441	2	A96759
35	66.5	8.7	769	2	H97033
36	66.5	8.7	914	1	JN0550
37	66.5	8.7	914	1	S07047
38	66.5	8.7	1140	2	F88349
39	66.5	8.7	1140	2	T20984
40	66	8.7	130	2	F95328
41	66	8.7	290	2	S41555
42	66	8.7	290	2	S69841
43	66	8.7	875	2	F70755
44	66	8.7	1058	2	T19282
45	66	8.7	1498	2	S53577

## ALIGNMENTS

## RESULT 1

S60178  
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy  
C:Species: Fusarium oxysporum  
C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000  
C:Accession: S60178  
R:Anaya, N.; Roncero, M.I.G.  
Mol. Gen. Genet. 249, 637-647, 1995  
A:Title: Skippy, a retrotransposon from the fungal plant pathogen Fusarium oxysporum  
A:Reference number: S60178; MID:96132549  
A:Accession: S60178  
A:Molecule type: DNA  
A:Residues: 1-853 <ANA>  
A:Cross-references: EMBL:L34658; NID:9510695; PIDN:AAA8790.1; PID:9510696  
A:Experimental source: retrotransposon skippy  
C:Genetics:  
A:Mobile element: retrotransposon skippy  
C:Keywords: polyprotein

Query Match 15.9%; Score 121; DB 2; Length 853;  
Best Local Similarity 34.5%; Pred. No. 0.00012;  
Matches 29; Conservative 12; Mismatches 31; Indels 12; Gaps 2;

OY 47 LRPAARRMRPIPF-----DETEDGDTDRLEPIVGTCTMYEDENTFSMDALKTFLIT 101  
DB 100 LASAANGRDPEVLKSPPEYFGTSPKPTFLTQSRATITVYPNQFRNDSAKVTMAG 159  
OY 102 RLTCPALQWVPIYRKESPLINDY 125  
DB 160 RLITKAQWF-----QPIINDY 176

## RESULT 2

T18347  
gag protein homolog, truncated - rice blast fungus magnaporthe gypsy retrotransposon  
C:Species: Magnaporthe grisea (rice blast fungus)  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T18347  
R:Farman, M.L.; Tosa, Y.; Nitta, N.; Leong, S.A.  
Mol. Gen. Genet. 251, 665-674, 1996  
A:Title: MAGT, a retrotransposon in the genome of the rice blast fungus Magnaporthe  
A:Reference number: Z18882; MUID:96335141  
A:Accession: T18347  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-457 <FAR>  
A:Cross-references: EMBL:L35053; NID:9522300; PID:9522301; PIDN:AAA33419.1  
C:Genetics:  
A:Gene: gag  
A:Mobile element: magnaporthe gypsy retrotransposon

Query Match

13.6%; Score 104; DB 2; Length 457;

serine/threonine p  
3'(2'),5'-bisphosph  
effector cell prot  
probable SRP fam11  
protein serine car  
alpha-glucosidase  
iodide peroxidase  
protein F15D4.7 (1  
hypothetical prote  
hypothetical prote  
Tya protein - yeas  
Tya protein - yeas  
hypothetical prote  
hypothetical prote  
Tya protein - yeas



Best Local Similarity 27.0%; Pred. No. 0.0038;  
Matches 24; Conservative 17; Mismatches 48; Indels 0; Gaps 0;

OY 49 PARRRRNPIPPETFDGDTLRPEIVOTCSMFEDNENFSDALKVFLIRLGPAL 108  
DB 152 PASARSERLPPDKPTGARSRLRRPATQIRKMTSNKDFPESRLIYIAGRLSGKAY 211

OY 109 OMVPIRKESPLINDYRGFLAEKRVFG 137  
DB 212 NLLPKMVGCTPQFGDTLQYLEARFG 240

## RESULT 3

gag polyprotein homolog - fungus (Cladosporium fulvum) retrotransposon Cfr-1

C:Species: Cladosporium fulvum  
C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 21-Jul-2000

C:Accession: S23569; S19849  
R:McHale, M.T.; Roberts, I.N.; Noble, S.M.; Beaumont, C.; Whitehead, M.P.; Seth, D.; Oik

Mol. Gen. Genet. 233, 337-347, 1992  
A:Title: Cfr-I: an LTR-retrotransposon in Cladosporium fulvum, a fungal pathogen of toma

A:Reference number: S23569; MID:92318865  
A:Accession: S23569

A:Molecule type: DNA  
A:Residues: 1-639 <MCH>

A:Cross-references: EMBL:Z11666; NID:q2562; PIDN:CAA77890.1; PID:q2563

A:Note: the authors translated the codon AOC for residue 55 as Ala, CAA for residue 451

C:Genetics:  
A:Mobile element: retrotransposon Cfr-1

C:Superfamily: Cladosporium fulvum probable gag polyprotein

Query Match 11.1%; Score 85; DB 2; Length 639;  
Best Local Similarity 30.2%; Pred. No. 0.63;  
Matches 26; Conservative 8; Mismatches 40; Indels 12; Gaps 2;

OY 61 PETFDSDTDLRPEIVOTCSYMEVDNENFSDALKVFLIRLGPALQWVPIRK-- 117  
DB 19 PEPFYGDGVAFDTWVSQMDYFLFNSMT---ENLKPIFATFLGRAGHWWKPLRKYL 75

OY 118 -----ESPLINDYRGFLAEKRVFG 137  
DB 76 SNGEDNADGVFKSYNHLKMAKSVFG 101

## RESULT 4

JC6323  
endonuclease (EC 3.1.21.-) - Bacillus globigii

C:Species: Bacillus globigii  
C:Date: 14-May-1998 #sequence\_revision 29-May-1998 #text\_change 24-Nov-1999

C:Accession: JC6323  
R:Anton, B.P.; Heiter, D.F.; Benner, J.S.; Hess, E.J.; Greenough, L.; Moran, L.S.; Slack

Gene 187, 19-27, 1997  
A:Title: Cloning and characterization of the BglII restriction-modification system reveal

A:Reference number: JC6323; MID:9725792  
A:Accession: JC6323

A:Molecule type: DNA  
A:Residues: 1-223 <ANT>

A:Cross-references: GB:U49842; NID:q1293617; PIDN:AAC45060.1; PID:q1293619

A:Experimental source: K08562  
C:Genetics:

A:Gene: bglIIR  
C:Superfamily: Bacillus globigii endonuclease

C:Keywords: hydrolase  
Query Match 10.2%; Score 78; DB 2; Length 223;  
Best Local Similarity 28.7%; Pred. No. 0.97;  
Matches 33; Conservative 16; Mismatches 36; Indels 30; Gaps 7;

OY 52 RRANPIPPETFD-----GDDRLRPEIVOTCSYMEVDNENF-----FSN 91  
DB 64 KHMKNIPIRKRFGLGIDIDGKRDLYE--VQFSNIPFLNNTVRSELEFKSNMID 121

OY 92 DALKVFLLIRLNG---PALQWVPIRKESPL--LNDYRGFLAEKRVFGMEED 141  
DB 122 EGMKVALITRK--GHMFPASSSLYYEAOQNOQLNSLAIFYVDFVPI-RLVGLIED 173

## RESULT 5

E87416  
monooxygenase, flavin-binding family [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: E87416  
R:Neiman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; J

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MID:21173698; PMID:11259647

A:Accession: E87416  
A>Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-499 <STO>

A:Cross-references: GB:AE005673; NID:q13422697; PIDN:AAK23329.1; GSPDB:GN00148

C:Genetics:  
A:Gene: CCI348

Query Match 10.2%; Score 77.5; DB 2; Length 499;  
Best Local Similarity 22.4%; Pred. No. 3;  
Matches 46; Conservative 21; Mismatches 63; Indels 75; Gaps 10;

OY 11 RRRISADPHATQNSARAGTMDGRVOLMKRL-----LAQPLRPAARRNPIRPF 61  
DB 252 RRRVLFQHEFTRAIESDITV--KAELEGVKMFGEEDFVAHFTPRYRWRORLAFV 309

OY 62 ETFGD-----FDRLPEF-----IVOTCSYMEVDNENF 89  
DB 310 P--DGLDFGVASGAKASVYDEICFTKGTLLSGETLDADYITATGPLSLVIGDIAF 367

OY 90 SND-----ALKVFLIRLNG-PALQWVPIRKESPLNDYNG-----FLAEK---- 133  
DB 368 EIDGQPLDFAKTYIRGMMFTGVPNLVWFGYFRASWTLRADLLGDFVCRLAHMERKGA 427

OY 134 -----RVFGMEDEDF 144  
DB 428 KQVEALRPEDDKMIGWIDPEDF 452

## RESULT 6

JE0163  
myelin expression factor-3 - mouse

N:Alternate names: MyEF-3  
C:Species: Mus musculus (house mouse)

C:Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999

C:Accession: JE0163  
R:Stephens, A.; Krynka, B.; Treliakova, A.; Haas, S.; Khalili, K.; Amin, S.

Biocem. Biophys. Res. Commun. 243, 295-301, 1998  
A:Title: MyEF-3, a developmentally controlled brain-derived nuclear protein which s

A:Reference number: JE0163; MID:98139908  
A:Accession: JE0163

A:Molecule type: mRNA  
A:Residues: 1-232 <STP>

A:Experimental source: brain  
C:Keywords: phosphoprotein

F:40-60/Domain: transmembrane #status predicted <TM>  
F:31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pre

F:31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #st

Query Match 10.1%; Score 77; DB 2; Length 232;  
Best Local Similarity 38.3%; Pred. No. 1.3;  
Matches 18; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

OY 61 PETFDGDTDLRPEIVOTCSYMEVDNENFSDALKVFLIRLGPAL 107  
DB 7 PEKFDGNFDMIGPFWYQCOLEFMEKSTROFSDVRINVCFTSLIGRA 53

## RESULT 7

T01324

hypothetical protein III\_4H10.1 - maize

C:Species: Zea mays (maize)

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 29-Oct-1999

C:Accession: T01324

R:Liaca, V.; Messing, J.

A:Description: Structure and organization of the 22-kDa alpha zeln gene cluster in Zea m

A:Reference number: Z14294

A:Accession: T01324

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-464 &lt;L1A&gt;

A:Cross-references: EMBL:AF031569; NID:g2832242; PIDN:AC01574.1; PID:g2832244

## Query Match

Best Local Similarity 20.8%; Score 74.5; DB 2; Length 464;

Matches 32; Conservative 17; Mismatches 60; Indels 45; Gaps 5;

Db 12 RLRSADPHATQNSAEARGTMDGRVQMLKALLAGPLRPAARMNPPI----- 59

Db 139 RRAGEARVSLERACERONIEGRNLDQDAEVVPOAPMGTRSOAGVPLACGCAALVDH 198

Db 60 -----FPETFGDGTDLRLPEFIYQTCSEMFVDENTFSNDALKYFLITRLTGP 106

Db 199 LRASMSKRFPHLPEKYDQTSN--PSEPIQV--YVATITAGCNTAMATYFHVAGSP 254

Db 107 ALQWVPIYIRKESPLNDYRGFLAEKRVFGWEE 140

Db 255 ARTW-----LNMISPG-----SIYSWEE 272

## RESULT 8

F70778

hypothetical protein RV2242 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: F70778

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

A:Conor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

A:Reference number: 537-544, 1998

A:Authors: Sgarbes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: F70778

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-414 &lt;COL&gt;

A:Cross-references: GB:Z70692; GB:AL123456; NID:g3261567; PIDN:CAA94663.1; PID:e235181;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV2242

## Query Match

Best Local Similarity 25.8%; Score 74; DB 2; Length 414;

Matches 24; Conservative 17; Mismatches 32; Indels 20; Gaps 4;

Db 8 MRFRRLSADPHATQNSAEARGTMDGRVQ--LMLKALLAGPLR-----AARRV----- 54

Db 141 LKYSNDLAFATATAYADAAEARGTMDSRMSAYDAVVRGDTGPELISRAALNMDTAP 200

Db 55 -----RNPFP--ETFGDGTDLRLPEFIYQTC 80

Db 201 ATVLVGTAPGPGNSGDSGRASQDVDRTA 233

## RESULT 9

G83758

oligoendopeptidase F BH0871 [Imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: G83758

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodure

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: G83758

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-589 &lt;STO&gt;

A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04590.1; GSI

C:Genetics:

A:Gene: BH0871

## Query Match

Best Local Similarity 24.2%; Score 73.5; DB 2; Length 589;

Matches 36; Conservative 17; Mismatches 49; Indels 47; Gaps 7;

Db 5 TLDNRFRRLSADPHATQNSAEARGTMDGRVQMLKALLAGPLRPAARMNPPI-----R 55

Db 129 TFSINERQAAAEKRLAPEQERKVAQAVDG-----YANMGEVYQAVGR 172

Db 56 NPFPETFGDGTDLRLPEFIYQTCSEMFVDENTFSNDALKYFLITRLTGPALQWVPIYI 115

Db 173 MEIPFE--DETRRLSGQL-----SNRLSSKRRKVRKASKVAEA--W----- 214

Db 116 RKESPL-----LNDYRGFLAEKRVFGWEE 140

Db 215 KKETPLFASTLNHLAGFLNLYRAGWMD 243

## RESULT 10

T08162

amylopullulanase (EC 3.2.1.-) - Pyrococcus furiosus

C:Species: Pyrococcus furiosus

C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999

C:Accession: T08162

R:Dong, G.; Vieille, C.; Zeikus, J.G.

A:Title: Cloning, sequencing, and expression of the gene encoding amylopullulanase

A:Reference number: Z16389; MUID:97438521

A:Accession: T08162

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-853 &lt;DON&gt;

A:Cross-references: EMBL:AF016588; NID:g2435436; PIDN:AAB71229.1; PID:g2435438

C:Genetics:

A:Gene: apu

C:Function:

A:Description: hydrolyzes alpha-1,6 and alpha-1,4 linkages in starch-related pol;

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

## Query Match

Best Local Similarity 22.7%; Score 73; DB 2; Length 853;

Matches 37; Conservative 16; Mismatches 52; Indels 58; Gaps 7;

Db 21 TQNSAEARGTMDGRVQMLKALLAGPLRPAARMNPPIFPETFGDGTDLRLPE----- 73

Db 145 TDENGNPIRDFWDRTELKDKMLA-----AKQKYN-LPLEBQKVAVTNEFTBODYDIA 198

Db 74 -----PIVOT-----CSYMFVDENTFSNDALKYFLITRL----- 103

Db 199 VLFNLAWIDYNYLIISPELKALYDKVDEGGYREDK--TVLYHQMWLNNTEFEHEKINL 257

Db 104 ---TGPALQWVPIYIRKESPLNDYRGFLAEKRVFGWEMED 143

Db 258 LLGNGVEVTVVYAHPIGPIILND-----FGWSDFD 289

## RESULT 11

T35628

probable penicillin-binding protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T35628

R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, June 1999

A:Reference number: 221584

A:Accession: T35628

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-485 <SEP>

A:Cross-references: EMBL:AL079356; PIDN:CAH45623.1; GSPDB:GN00070; SCOEDB:SC669.32

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC669.32

Query Match 9.4%; Score 71.5; DB 2; Length 485;

Best Local Similarity 26.6%; Pred. No. 13;

Matches 33; Conservative 15; Mismatches 43; Indels 33; Gaps 5;

Db 12 RRLSDPHATQNSAEAR---GTMDGRVQLMKALLAGPLRPAARRRNPPIPFEDGD 67

193 KILNADPDKPMLNRAVSKRTYPPGCTKVVYTAALRAGVIRDLADPTSPDP--TLPGI 250

Qy 68 TDLRP-----EFIVQCSYMF---VDENT-----FSNDALKYTFPI 100

Db 251 RRLTFNEADGCRNASLREAFEMSCNTVFALGVDCVRCMTAMAEAFNFDGLRVDFPV 310

Qy 101 TRLT 104

Db 311 ARST 314

RESULT 12

T21194

hypothetical protein F21D5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T21194

R:Berts, M. submitted to the EMBL Data Library, September 1995

A:Reference number: 219389

A:Accession: T21194

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-550 <WIL>

A:Cross-references: EMBL:254271; PIDN:CAA91032.1; GSPDB:GN00022; CESP:F21D5.1

A:Experimental source: Clone F21D5

C:Genetics:

A:Gene: CESP:F21D5.1

A:Map position: 4

A:Introns: 33/1; 97/3; 213/2; 329/3; 409/2; 512/2

C:Superfamily: Caenorhabditis elegans hypothetical protein F21D5.1

Query Match 9.4%; Score 71.5; DB 2; Length 550;

Best Local Similarity 29.7%; Pred. No. 15;

Matches 38; Conservative 13; Mismatches 40; Indels 37; Gaps 8;

Qy 9 RFRRLSADPHATQ---RSAAE--RGTMGRVQLMKALLAGPLRPAARRRNPPIPFPE 62

Db 230 RFRNLRLPSSLLVEFRNESELNCGADPFVKISOKLPAN-FSPFAE-----PKCA 283

Qy 63 TFDGDTDLRLPEFIVQCSYFVDENTFSNDAL-----KVFLLTRLGLPALQWVPIYIRK 117

Db 284 SFGDADRLMYFRAK-----ASENSNSDAELFDGXIIVLTI-----VYIIRE 326

Qy 118 ESPLINDY 125

Db 327 Q---LMDY 331

RESULT 13

D82550

conserved hypothetical protein XF2493 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: D82550

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide S

tructure 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82517; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-180 <SIM>

A:Cross-references: GB:AE004057; GB:AE003849; NID:99107690; PIDN:AAF65291.1; GSPDB:

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga,

Bitones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanli, A.P.; Ferreira, A.O.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.;

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.;

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Marti

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki,

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeira,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanceli, R.V.; Se

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Si

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2493

C:Superfamily: Pseudomonas aeruginosa hypothetical protein PRF10

Query Match 9.2%; Score 70.5; DB 2; Length 180;

Best Local Similarity 26.8%; Pred. No. 4.7;

Matches 33; Conservative 16; Mismatches 39; Indels 35; Gaps 7;

Qy 4 PTLMRFRRL--SADP--HATQNSAEARGTMDGRVQLMKALLAGPLRPAARRRNPPI 58

Db 68 PAL-LRFQARLVGHRDPFTTHQARLAAALGTW-----LYQRRRG-1 109

Qy 59 P-----FPETPGDTRLRPEFIVQCSYFVDENTFSNDALKVTPL---ITRLTGP 106

Db 110 PTDPGRVIALYPDEFNAELDRVYVWQOPVFLGNSAMTHAGTLPHTLMTSFTQTGP 169

Qy 107 ALQ 109

Db 170 SFQ 172

RESULT 14

S59771

hypothetical protein YPR106w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein P8283.9

C:Species: Saccharomyces cerevisiae

C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 06-Feb-1998

C:Accession: S59771

R:Nelson, J. submitted to the EMBL Data Library, July 1995

A:Description: The sequence of S. cerevisiae cosmid 8283.

A:Reference number: S59764

A:Accession: S59771

A:Molecule type: DNA

A:Residues: 1-443 <NEL>

A:Cross-references: EMBL:U02445; NID:9914969; PID:9914977; MIPS:YPR106w

C:Genetics:

A:Gene: SGD:ISRI

A:Cross-references: SGD:S0006310; MIPS:YPR106w

A:Map position: 16R

Query Match 9.2%; Score 70.5; DB 2; Length 443;

Best Local Similarity 22.7%; Pred. No. 14;

Matches 22; Conservative 18; Mismatches 32; Indels 25; Gaps 4;



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 02:39:39 ; Search time 11 Seconds  
(without alignments)  
506.874 Million cell updates/sec

Title: US-09-673-400a-38

Perfect score: 763  
Sequence: 1 ARAPFLDMRFRRLSADPHA.....YRGLAEMKRVGWEDEDF 144

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	335	43.9	146	1	LDOL_HUMAN
2	74	9.7	414	1	YMG2_MYCTU
3	74	9.7	970	1	NAC1_MOUSE
4	69	9.0	348	1	RM03_HUMAN
5	69	9.0	1487	1	ICP4_HSVK
6	68.5	9.0	1515	1	YCF1_YEAST
7	68	8.9	514	1	OAS1_HUMAN
8	67.5	8.8	316	1	CYR1_STRCO
9	67.5	8.8	971	1	NAC1_RAT
10	66.5	8.7	430	1	FTSY_MYCLE
11	66.5	8.7	914	1	PERT_MOUSE
12	66.5	8.7	914	1	PERT_RAT
13	66	8.7	488	1	SYK_MYCFE
14	66	8.7	875	1	YCF8_MYCTU
15	66	8.7	989	1	RPOC_LEUME
16	66	8.7	159	1	SEC7_YEAST
17	65.5	8.6	159	1	LATX_PAPSO
18	65.5	8.6	822	1	HIFA_MOUSE
19	65.5	8.6	4823	1	BIR6_HUMAN
20	65	8.5	369	1	HAIR_MOUSE
21	65	8.5	819	1	ION_CHLMU
22	65	8.5	863	1	AMPN_CAVCR
23	65	8.5	1041	1	ATNA_DROME
24	65	8.5	1487	1	ICP4_HSVK
25	64.5	8.5	906	1	FOX2_CANTR
26	64.5	8.5	910	1	IMB2_SCHPO
27	64.5	8.5	4639	1	DYRC_DROME
28	64	8.4	614	1	FMRI_MOUSE
29	64	8.4	819	1	ION_CHLMU
30	64	8.4	1597	1	LRIL_YEAST
31	63.5	8.3	139	1	IL11_HUMAN
32	63.5	8.3	395	1	IF5_SCHPO
33	63.5	8.3	416	1	G3PA_GRAVE

34	63.5	8.3	630	1	TNFB_STAAT
35	63.5	8.3	874	1	UIS2_EBV
36	63.5	8.3	1698	1	Y076_HUMAN
37	63	8.3	619	1	YAL6_CHLPN
38	63	8.3	724	1	MALO_MYCTU
39	63	8.3	819	1	ION_CHLPN
40	63	8.3	973	1	NAC1_HUMAN
41	63	8.3	1005	1	YCF1_OENB
42	62.5	8.2	472	1	SACH_BACAM
43	62.5	8.2	671	1	RIP_HUMAN
44	62	8.1	303	1	YER8_METJA
45	62	8.1	348	1	RM03_MOUSE

## ALIGNMENTS

RESULT 1	ID	LDOL_HUMAN	STANDARD:	PRT:	146 AA.
AC	095751:				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	LDOL1 protein (Leucine zipper protein down-regulated in cancer cells).				
GN	LDOL1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal Brain;				
RX	MEDLINE=99330357; PubMed=10403563;				
RA	Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T.,				
RA	Yamaguchi K.;				
RT	Identification of a novel gene, LDOL1, down-regulated in cancer cell				
RT	lines.;				
RL	Cancer Lett. 140:227-234(1999).				
CC	-1- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN THE DEVELOPMENT AND/OR				
CC	PROGRESSION OF SOME CANCERS.				
CC	-1- SUBCELLULAR LOCATION: Nuclear.				
CC	-1- TISSUE SPECIFICITY: OBVIOUSLY EXPRESSED WITH HIGH LEVELS IN				
CC	BRAIN ANT THYROID AND LOW EXPRESSION IN PLACENTA, LIVER AND				
CC	LEUCOCYTES. EXPRESSED AS WELL IN SIX OF THE SEVEN HUMAN BREAST				
CC	CANCER CELL LINES EXAMINED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL; AB019527; BAA34364.1; -				
DR	Nuclear protein.				
KW	DOMAIN 132 143				
FT	SEQUENCE 146 AA; 16968 MW; 3D02813B2DE52DBE CRC64;				
SO	SEQUENCE				
Query Match	43.96; Score 335; DB 1; Length 146;				
Best Local Similarity	59.68; Pred. No. 6.5e-30;				
Matches	62; Conservative 17; Mismatches 21; Indels 4; Gaps 1;				
QY	40 KALLAGLRPARWRNRDIPPEFDDGTDLPEFYOTCSYMDVNTSNDALKYFL 99				
DB	37 RASLRLQVYRPPS---CPVPPETFGNSSSLPEFYQTASVYLVNRRFCNKAFL 92				
QY	100 ITRLTGPALOWIVPIRKESPLNDYRGFLAEMKRVGWEDEDF 143				
DB	93 ISLTGRAEWVYPIIENDSPILGDYRAFLDEMKQRCGWDDDD 136				

```

RESULT 2
YM42_MYCTU STANDARD; PRT; 414 AA.
ID YM42_MYCTU
AC 010523;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 44.6 kDa protein RV2242.
GN RV2242 OR MT2302 OR MTCY427.23.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock R., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hovnsby T., Jagsels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CDAR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL: Z70692; CA94663.1;
DR EMBL: AE007074; AAK46586.1;
DR TIGR: MT2302;
DR Tuberculist: RV2242;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 414 AA; 44637 MW; F454D3397711F73 CRC64;

Query Match 9.7%; Score 74; DB 1; Length 414;
Best Local Similarity 25.8%; Pred. No. 1.5;
Matches 24; Conservative 17; Mismatches 32; Indels 20; Gaps 4;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium/calcium exchanger 1 precursor (Na(+)/Ca(2+)-exchange protein
DE 1).
GN SLC8A1 OR NCX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=96250070; PubMed=8659820;
RA Kim I., Lee C.O.;
RT "Cloning of the mouse cardiac Na(+)-Ca2+ exchanger and functional
RT expression in Xenopus oocytes."
RL Ann. N.Y. Acad. Sci. 779:126-128(1996).
CC -1- FUNCTION: RAPIDLY TRANSPORTS CA2+ DURING EXCITATION-CONTRACTION
CC COUPLING. CA(2+) IS EXTRUDED FROM THE CELL DURING RELAXATION
CC SO AS TO PREVENT OVERLOADING OF INTRACELLULAR STORES.
CC -1- ENZYME REGULATION: BY ATP...
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane.
CC -1- TISSUE SPECIFICITY: CARDIAC SARCOLEMA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL: U70033; AAB46708.1;
DR MGD: MGI:107956; SLC8a1.
DR InterPro: IPR003644; Calx_beta.
DR InterPro: IPR002613; Na_Ca_Ex.
DR Pfam: PF01699; Na_Ca_Ex_2.
DR PRINTS: PR01259; NACAEXCHNR.
DR SMART: SM00237; Calx_beta; 2.
KW Transport; Symport; Transmembrane; Glycoprotein; Phosphorylation;
KW Signal; Calmodulin-binding; Calcium transport; Repeat.
CC
CC SIGNAL 1 32
CC CHAIN 33 970
CC DOMAIN 33 71
CC TRANSMEM 72 93
CC DOMAIN 94 133
CC TRANSMEM 134 155
CC DOMAIN 156 167
CC TRANSMEM 168 188
CC DOMAIN 189 199
CC TRANSMEM 200 222
CC DOMAIN 223 225
CC TRANSMEM 226 249
CC DOMAIN 250 769
CC TRANSMEM 770 789
CC DOMAIN 790 796
CC TRANSMEM 797 819
CC DOMAIN 820 821
CC TRANSMEM 822 840
CC DOMAIN 841 871
CC TRANSMEM 872 882
CC DOMAIN 893 903
CC TRANSMEM 904 924
CC DOMAIN 925 941
CC TRANSMEM 942 958
CC DOMAIN 959 970
CC TRANSMEM 970 270
CC DOMAIN 251 270
CC REPEAT 158 178
CC REPEAT 407 478
CC REPEAT 539 609
CC REPEAT 839 875
CC DOMAIN 236 239

```

Query Match 9.7% Score 74; DB 1; Length 970;  
Best Local Similarity 27.0%; Pred. No. 4.4;  
Matches 34; Conservative 20; Mismatches 40; Indels 32; Gaps 7;

FT DOMAIN 689 692 POLY-GLT.  
FT DOMAIN 756 760 POLY-ASP.  
FT MOD\_RES 389 389 PHOSPHORYLATION (POTENTIAL).  
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 970 AA: 108035 MW: 15500BD07F2B6602 CRC64;

Query Match 9.7% Score 74; DB 1; Length 970;  
Best Local Similarity 27.0%; Pred. No. 4.4;  
Matches 34; Conservative 20; Mismatches 40; Indels 32; Gaps 7;

DB 547 SESIGIMEYKV-----LRTSGARGNVILPY-KTIEGTARGGCEDEFCGSEFQ 595  
OY 26 AEARCTMDGRVQMLKALLAGPLRPARARWRNPPEPEFDG-----DTRLPFR- 74  
DB 596 NDEIVKTSVQVDEEYERKN--TFEL-ELGEPRL--VENSEKALLNLNLSGFTLT 648  
OY 132 MKRVFG 137  
DB 649 GKEMTG 654

RESULT 4  
RM03\_HUMAN STANDARD: PRT: 348 AA.

AC P09001.  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Mitochondrial 60S ribosomal protein L3.  
GN MRPL3 OR MR13.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP -SEQUENCE FROM N.A.  
RX MEDLINE=88067705; PubMed=2891103;  
RA Ou J.-H., Yen T.S.-B., Wang Y.-F., Kam W.K., Rutter W.;  
RT "Cloning and characterization of a human ribosomal protein gene with  
RT enhanced expression in fetal and neoplastic cells.";  
RL Nucleic Acids Res. 15:8919-8934(1987).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X06323; CAA29639.1; -  
CC PIR: S01607; R5HUL3.  
DR InterPro: IPR000597; Ribosomal\_L3.  
DR Pfam: PF00297; Ribosomal\_L3; 1.  
DR PRODOM: PD001374; Ribosomal\_L3; 1.  
DR PROSITE: PS00474; RIBOSOMAL\_L3; 1.  
KW Ribosomal protein; Mitochondrion.  
SQ SEQUENCE 348 AA: 38632 MW: 6D6B5BD6BD72E32 CRC64;

Query Match 9.0% Score 69; DB 1; Length 348;  
Best Local Similarity 29.1%; Pred. No. 4.4;  
Matches 23; Conservative 10; Mismatches 36; Indels 10; Gaps 3;

OY 21 TQNSAARCTMDGRVQMLKALLAGPLRPARARWRNPPEPEFDG-DGDTLRLEFIVQTC 79  
DB 278 TKHNIIVYNGSVPGHKNCILVAKDSKL-PAYKDLGKNLPPEYTFDGDDEELPDDL----- 332

OY 80 SYMFVDENTSNDALAKYTF 98  
DB 333 -----YDENWQPGAPPSITF 347

RESULT 5  
ICP4\_HSVK STANDARD: PRT: 1487 AA.

ID ICP4\_HSVK  
AC I7473;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Trans-acting transcriptional protein ICP4 (155 kDa immediate-early  
DE protein).  
GN IE OR 64.  
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OX Alphaherpesvirinae; Varicellovirus.  
RN [1]  
RP -SEQUENCE FROM N.A.  
RX MEDLINE=89370304; PubMed=2549711;  
RA Grundy F.J., Baumann R.P., O'Callaghan D.J.;  
RT "DNA sequence and comparative analyses of the equine herpesvirus type  
RT 1 immediate early gene.";  
RL Virology 172:223-236(1989).  
RN [2]  
RP -SEQUENCE OF 1432-1487 FROM N.A.  
RX MEDLINE=90064773; PubMed=2555546;  
RA Hartly R.N., Colle C.F. III, Grundy F.J., O'Callaghan D.J.;  
RT "Mapping the terminal and intron of the spliced immediate-early  
RT transcript of equine herpesvirus 1.";  
RL J. Virol. 63:5101-5110(1989).  
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE  
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING  
CC OTHER VIRAL GENES, AND ADJUSTING ITS OWN SYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.  
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF  
CC PHOSPHORYLATION.  
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: J04366; AAA45089.1; -  
CC DR EMBL: M30498; AAA66554.1; -  
CC PIR: A33764; EDBEEL.  
DR HSSP: P04002; IWEA.  
KW Early protein; Transcription regulation; Trans-acting factor;  
KW DNA-binding; Phosphorylation; Nuclear protein.  
FT DOMAIN 181 213 SER-RICH.  
FT DOMAIN 922 931 ARG/LYS-RICH (BASIC).  
SQ SEQUENCE 1487 AA: 154716 MW: 044E39A570608A6B CRC64;

Query Match 9.0% Score 69; DB 1; Length 1487;  
Best Local Similarity 24.6%; Pred. No. 26;  
Matches 33; Conservative 15; Mismatches 48; Indels 38; Gaps 5;

OY 10 FRRRLSADPHATQRSARAG--TWDGAVQMLKALLAGPLRPARARWRNPPEPEFDGD 67  
DB 1035 WRPALTEPPALATIAACSGPRARD-ARGELAAAGPLRRRA-AMHQLPDPED----- 1088  
OY 68 TDRLEPFIIVQCSYMFVDENTSNDALAKYTFILRLTGPALQWVPIYR-----KESPLN 123  
DB 1089 -----VKVVLVSLQEDDLGLGAPASPGSGSRRPELMS 1122  
OY 124 DYRGFLAMKRVFG 137

Db 1123 DLKGLSALLAALG 1136

RESULT 6

YCFL\_YEAST STANDARD: PRT: 1515 AA.

ID YCFL\_YEAST 003905;

AC P39109; 003905; (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 40, Last annotation update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Metal resistance protein YCFL (Yeast cadmium factor 1).

GN YCFL OR YDR135C OR YD9302.11C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H9;

RX MEDLINE=94357936; PubMed=7521334;

RA Sczyzka M.S., Memme J.A., Moye-Rowley S.W., Thiele D.J.;

RT "A yeast metal resistance protein similar to human cystic fibrosis transmembrane conductance regulator (CFTR) and multidrug resistance-associated protein."

RT J. Biol. Chem. 269:22853-22857(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=SZ88C / AB972;

RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;

RT Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.

RN [3]

RP FUNCTION.

RX MEDLINE=20253522; PubMed=10790694;

RA Petrovic S., Pascolo L., Gallo R., Cupelli F., Ostrow J.D., Goffeau A., Tiribelli C., Brusch C.V.;

RT "The products of YCFL and YL1015w (BPT1) cooperate for the ATP-dependent vacuolar transport of unconjugated bilirubin in Saccharomyces cerevisiae."

RT Yeast 16:561-571(2000).

CC -1- FUNCTION: COOPERATES FOR THE ATP-DEPENDENT VACUOLAR TRANSPORT OF BILIRUBIN.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: L35237; AAA50353.1; -

DR EMBL: Z48179; CA88217.1; -

DR HSP: P13569; INED.

DR SGD: S0002542; YCFL.

DR InterPro: IPR003593; AAA.

DR InterPro: IPR001140; ABC\_transporter\_tmam.

DR InterPro: IPR003439; ABC\_transporter.

DR Pfam: PF00664; ABC\_membrane\_2.

DR Pfam: PF00605; ABC\_tran\_2.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC\_TRANSPORTER; 2.

DR ATP-binding; Transmembrane; Glycoprotein; Transport;

KW Cadmium resistance; Phosphorylation.

FT DOMAIN 1 32 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 33 53 1 (BY SIMILARITY).

FT DOMAIN 54 73 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 74 94 2 (BY SIMILARITY).

FT DOMAIN 95 99 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 100 120 3 (BY SIMILARITY).

FT DOMAIN 121 130 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 131 151 4 (BY SIMILARITY).

FT DOMAIN 152 169 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 170 190 5 (BY SIMILARITY).

FT DOMAIN 191 278 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 279 299 6 (BY SIMILARITY).

FT DOMAIN 300 345 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 346 366 7 (BY SIMILARITY).

FT DOMAIN 367 422 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 423 443 8 (BY SIMILARITY).

FT DOMAIN 444 467 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 468 530 9 (BY SIMILARITY).

FT DOMAIN 531 551 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 552 572 10 (BY SIMILARITY).

FT DOMAIN 573 593 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 594 943 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 944 964 12 (BY SIMILARITY).

FT DOMAIN 965 1001 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1002 1023 13 (BY SIMILARITY).

FT DOMAIN 1024 1066 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 1067 1087 14 (BY SIMILARITY).

FT DOMAIN 1088 1088 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1089 1109 15 (BY SIMILARITY).

FT DOMAIN 1110 1180 CYTOPLASMIC (BY SIMILARITY).

FT TRANSMEM 1181 1201 16 (BY SIMILARITY).

FT DOMAIN 1202 1205 EXTRACELLULAR (BY SIMILARITY).

FT TRANSMEM 1206 1226 17 (BY SIMILARITY).

FT DOMAIN 1227 1515 CYTOPLASMIC (BY SIMILARITY).

FT NP\_BIND 663 670 ATP (POTENTIAL).

FT NP\_BIND 856 863 ATP (POTENTIAL).

FT NP\_BIND 1306 1313 ATP (POTENTIAL).

FT MOD\_RES 908 908 PHOSPHORYLATION (BY CAPK) (POTENTIAL).

FT MUTAGEN 713 713 MISSING: LOSS OF FUNCTION.

FT MUTAGEN 908 908 S->A: LOSS OF FUNCTION.

FT CONFLICT 680 680 L->R (IN REF. 1).

SO SEQUENCE 1515 AA; 171120 MW; 30F92FEDBAF60431 CRC64;

Query Match 9.0%; Score 68.5; DB 1; Length 1515;

Best Local Similarity 23.4%; Pred. No. 30;

Matches 29; Conservative 19; Mismatches 35; Indels 41; Gaps 5;

OY 55 RNP1PPEP-----FDGDTDRLEPIVOTCSYFVDENFNSDALKVTEFLITPLNG 105

DB 1035 RAMPTEFETPIGRILNRPSNDLYKVDALGRFSQEFV-----NAVVTITIVYICA 1087

OY 106 PALOW---VIP-----YIR-----KESPLNDYRCGLAEKRRVSGWEE 140

DB 1088 TTMQCFIIFILPSVFIYVQOYVILRSRELRDSTIRSPYISHFOTLGLATVNGYSQ 1147

OY 141 DEDF 144

DB 1148 QKRPF 1151

RESULT 7

OASL\_HUMAN STANDARD: PRT: 514 AA.

ID OASL\_HUMAN 015646; 075686; Q9Y6K7; Q9Y6K6;

AC 015646; 075686; Q9Y6K7; Q9Y6K6; (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE 59 kDa 2'-5'-oligoadenylate synthetase like protein (p59 OASL)

DE (p59OASL) (Thyroid receptor interacting protein 14) (TRIP14).

GN OASL OR TRIP14.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM P56).

RX MEDLINE=98391734; PubMed=9722630;

RA Hartmann R., Olsen H.S., Wilder S., Joergensen R., Justesen J.;

RT "p59OASL, a 2'-5' oligoadenylate synthetase like protein: a novel human gene related to the 2'-5' oligoadenylate synthetase family."



RL Nucleic Acids Res. 26:4121-4127(1998).  
 RN [12]  
 RP SEQUENCE FROM N.A. (ISOFORMS P56 AND P30).  
 RC TISSUE-Monocytes;  
 RX MEDLINE-99041549; PubMed-9826176;  
 RA Reboulilat D., Marie I., Hovanessian A.G.;  
 RT "Molecular cloning and characterization of two related and interferon-  
 induced 56-kDa and 30-kDa proteins highly similar to 2'-5'-  
 oligoadenylate synthetase."  
 RT Eur. J. Biochem. 257:319-330(1998).  
 RL [13]  
 RP SEQUENCE FROM N.A.  
 RA White S.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 260-416 FROM N.A. (ISOFORM P56).  
 RX MEDLINE-95295737; PubMed-7776974;  
 RA Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;  
 RT "Two classes of proteins dependent on either the presence or absence  
 of thyroid hormone for interaction with the thyroid hormone  
 receptor."  
 RT Mol. Endocrinol. 9:243-254(1995).  
 RL [1]  
 CC -1- FUNCTION: DOES NOT HAVE 2'-5'-OAS ACTIVITY, BUT BINDS DOUBLE-  
 STRANDED RNA AND DNA.  
 CC -1- SUBUNIT: SPECIFICALLY INTERACTS WITH THE LIGAND BINDING DOMAIN OF  
 THE THYROID RECEPTOR (TR). TRIP14 DOES NOT REQUIRE THE PRESENCE OF  
 THYROID HORMONE FOR ITS INTERACTION.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: P56 (SHOWN HERE) AND P30; MAY BE  
 PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH THE HIGHEST  
 LEVELS IN PRIMARY BLOOD LEUKOCYTES AND OTHER HEMATOPOIETIC SYSTEM  
 TISSUES, COLON, STOMACH AND TO SOME EXTENT IN TESTIS.  
 CC -1- INDUCTION: BY INTERFERONS.  
 CC -1- SIMILARITY: BELONGS TO THE 2'-5A SYNTHETASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 UNIQUTIN-LIKE DOMAINS.  
 CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 386  
 TO 416 DUE TO A FRAMESHIFT.  
 CC -1- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF MOUSE OASL.  
 CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT CONFLICT 95 113  
 FT CONFLICT 223 223  
 FT CONFLICT 317 317  
 FT CONFLICT 321 321  
 FT CONFLICT 324 324  
 FT CONFLICT 341 342  
 FT CONFLICT 445 445  
 SQ SEQUENCE 514 AA: 59226 MW: 4888655D9EA003E CRC64;  
 Query Match 8.9%; Score 68; DB 1; Length 514;  
 Best Local Similarity 24.3%; Pred. No. 9.2;  
 Matches 37; Conservative 24; Mismatches 55; Indels 36; Gaps 10;  
 QY 7 DMFRRRLSADPHATONSASARCTMDG-RVQIMKALLA-GILRPAARMRNPPIPPETP 64  
 DB 125 DLRMGEV--PDALV-FTIOTRGTAIPVTYIVAVRALGSPSP-----NSQPPPEVY 174  
 QY 65 DGDTRLPEVIVTQ-----SYMFVDTEFSDALKVFTLRLTGLPALOWIPIYR 116  
 DB 175 VS-----LIRACGGPGNFCPSFSELQRFVNRHPTKLSL-RL--VKHWYQOYVK 222  
 QY 117 KESPLNDYRGFLAEMKRVFGM---EDEDEF 144  
 DB 223 ARSFRANLPPVLALELLTITAMENGTEDENF 254  
 RESULT 8  
 CYRK\_STRCO STANDARD; PRT; 316 AA.  
 ID 002055;  
 AC 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Actinorhodin polyketide synthase bifunctional cyclase/dehydratase  
 DE (actI ORF4) (ACTVIT).  
 GN SCBAC2861.16.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-A3(2) / M145;  
 RX MEDLINE-92406871; PubMed-1527048;  
 RA Fernandez-Moreno M.A., Martinez E., Boto L., Hopwood D.A.,  
 Malpartida F.;  
 RT "Nucleotide sequence and deduced functions of a set of cotranscribed  
 RT genes of Streptomyces coelicolor A3(2). Including the polyketide  
 RT synthase for the antibiotic actinorhodin."  
 RL J. Biol. Chem. 267:19278-19290(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2).  
 RA Warren T., Harris D., Cerdano A.M., Parkhill J., Barrell B.G.,  
 Rajandream M.A.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: IS NEEDED FOR CORRECT CYCLIZATION OF THE OLIGOPEPTIDE  
 LEADING TO ISOCROMANQUINONE FORMATION.  
 CC -1- PATHWAY: BIOSYNTHESIS OF POLYKETIDE ANTIBIOTIC ACTINORHODIN.  
 CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X63449; CA45046.1; -  
 DR EMBL: AL593842; CA44203.1; -  
 DR PIR: S25843; S25843.  
 KW Antibiotic biosynthesis.



FT CONFLICT 402 402 P -> A (IN REF. 1).  
SQ SEQUENCE 971 AA; 108184 MW; EC456CE3AC6A69 CRC64;  
Query Match  
Best Local Similarity 27.6%; Pred. No. 23;  
Matches 35; Conservative 21; Mismatches 38; Indels 33; Gaps 8;  
OY 26 AEARGTGDRVQMLKALLAGPLRPARARRNPDPPEFTEDG-----DTDRLPPEF-74  
DB 547 SESIGIMEVKV-----LRTSGARGNVITPY-KTTEGTARGGDEPDTCGELEFQ 595  
OY 75 ---IVQTSYMFVDENTSNDALKTFLITRLTGALQWVPIRKESPLNDYGF-LA 130  
DB 596 NDEIVKTIISVYVIDEEYEKNK---TFPI-EIGEPRL---VEMSEKALLNELGGFTLT 648  
OY 131 EMKRVFG 137  
DB 649 EGKKTG 655  
RESULT 10  
FTSY\_MYCLE  
ID FTSY\_MYCLE STANDARD: PRT: 430 AA.  
AC 033010;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cell division protein ftsy homolog.  
GN FTSY OR ML1628 OR MLCB250.02.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID:1769;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-TN:  
RC MEDLINE=21128732; PubMed=11234002;  
RX Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Houtby L., Hornsby T., Jagels K., Jactroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Steeger K., Simon S., Simmonds M., Skellon J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
CC -1- FUNCTION: FUNCTIONAL HOMOLOGY OF A SUBSET OF PROTEINS AT THE  
THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE  
MEMBRANE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: 297369; CAB10596.1;  
DR EMBL: AL583922; GAC30579.1;  
DR HSSP: P10121; 1FTS.  
DR Leptoma; ML1628;  
DR InterPro: IPR003593; AAA.  
DR InterPro: IPR001687; ATP-GTP\_A.  
DR InterPro: IPR000897; SRP54.  
DR Pfam: PF00448; SRP54; 1.  
DR Pfam: PF02881; SRP54.N; 1.  
DR ProDom: PD000819; SRP54; 1.  
DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00300; SRP54; 1.  
KW Signal recognition particle; GTP-binding; RNA-binding; Membrane;  
KW Cell division; Complete proteome.  
FT NP\_BIND 238 245 GTP (BY SIMILARITY).  
FT NP\_BIND 320 324 GTP (BY SIMILARITY).  
FT NP\_BIND 382 385 GTP (BY SIMILARITY).  
SQ SEQUENCE 430 AA; 45354 MW; E9F78589AAB2C598 CRC64;  
Query Match  
Best Local Similarity 24.5%; Pred. No. 11;  
Matches 40; Conservative 19; Mismatches 57; Indels 47; Gaps 6;  
OY 8 MRRERR-----LSADPHATQNSAEKRGTMGRVQMLKALLAGPLRPARARRNPDP 59  
DB 26 VYRRRRRLSLPSSLSGSGASDRS---GSYVSSGTFPSOTALPVQADRI----- 74  
OY 60 FPEFDGPTDRLEPFTVQCSYMFVDENTSNDALKTFLITRLTGALQWVPIRKES 119  
DB 75 -----DTGELPVGDDATVPRDSRHTISDVLPESSELITSPDEP--EAAVPHIDATV 125  
OY 120 P---LLNDYRGFLAEKRVFG-----WEDED 143  
DB 126 PREGRLDRLRGRLARTQNAFGRSILGLVGDLEDSDMQELSD 168  
RESULT 11  
PERT\_MOUSE  
ID PERT\_MOUSE STANDARD: PRT: 914 AA.  
AC P35419;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Thyroid peroxidase precursor (EC 1.11.1.8) (TPO).  
GN TPO.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID:10090;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-C57BL/6; TISSUE=Thyroid;  
RX MEDLINE=93154601; PubMed=7916704;  
RA Kotani T., Umeki K., Yamamoto I., Takeuchi M., Takeuchi S.,  
RA Nakayama T., Ohtaki S.;  
RT "Nucleotide sequence of the cDNA encoding mouse thyroid peroxidase.";  
RL Gene 123:289-290(1993).  
CC -1- FUNCTION: IODINATION AND COUPLING OF THE HORMONAL TYROSINES  
IN THYROGLOBULIN TO YIELD THE THYROID HORMONES T(3) AND T(4).  
CC -1- CATALYTIC ACTIVITY: Iodide + H(2)O(2) = Iodine + 2 H(2)O.  
CC -1- COFACTOR: HEME (PROTOPORPHYRIN IX).  
CC -1- PATHWAY: THYROID HORMONE BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Membrane.  
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X60703; CAA43114.1;  
DR PIR: JN0550; JN0550.  
DR HSSP: P05164; 1CXP.  
DR MGD: MGT:98813; Tpo.  
DR InterPro: IPR002007; Anim\_peroxidase.  
DR InterPro: IPR000157; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF-Ca.  
DR InterPro: IPR001536; Peroxidase\_3.  
DR InterPro: IPR002016; Peroxidase.

DR Interpro: IPR000436; Sushi\_SCR.CCP.  
 DR Pfam: PF03098; An\_peroxidase; 1.  
 DR Pfam: PF00008; EGF; 1.  
 DR Pfam: PF00084; Sushi; 1.  
 DR PRINTS; PR00457; ANPEROXIDASE.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR Oxidoreductase; Peroxidase; Heme; Transmembrane; Glycoprotein;  
 KW Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 914 THYROID PEROXIDASE.  
 FT ACT\_SITE 233 233 DISTAL HISTIDINE (POTENTIAL).  
 FT ACT\_SITE 384 384 DISTAL ARGININE (POTENTIAL).  
 FT ACT\_SITE 482 482 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)  
 FT ACT\_SITE (POTENTIAL).  
 FT TRANSMEM 835 859 POTENTIAL.  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 914 AA; 101342 MW; 59569A0B71F3DD01 CRC64;

Query Match 8.7%; Score 66.5; DB 1; Length 914;  
 Best Local Similarity 30.9%; Pred. No. 27;  
 Matches 21; Conservative 12; Mismatches 32; Indels 3; Gaps 2;

OY 61 PETFGDRLPEFIVOCYMFVDENTFSNDALKTFL-ITRLTGPALQWVPIYIRKE 118  
 DB 80 PESTSGAISRAELMETSIQWKREQSOFSTDALSDILATIANLSG-CLPFLMLPPRCDD 138  
 OY 119 SPLNDYR 126  
 DB 139 TCLANKYR 146

## RESULT 12

ID PERT\_RAT STANDARD; PRT; 914 AA.  
 AC P14650;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Thyroid peroxidase precursor (EC 1.11.1.8) (TPO).  
 GN TPO.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90045972; PubMed-2813071;  
 RA Dewahl M., Seto P., Rapoport B.;  
 RT "Complete nucleotide sequence of the cDNA for thyroid peroxidase in  
 FRTL5 rat thyroid cells."  
 RL Nucleic Acids Res. 17:8380-8380(1989).  
 RN [2]  
 RP SEQUENCE OF 145-914 FROM N.A.  
 RX MEDLINE-90114171; PubMed-2691880;  
 RA Iozaki O., Kohn L.D., Kozak C.A., Kimura S.;  
 RT "Thyroid peroxidase: rat cDNA sequence, chromosomal localization in  
 mouse, and regulation of gene expression by comparison to  
 thyroglobulin in rat FRTL-5 cells."  
 RL Mol. Endocrinol. 3:1681-1692(1989).  
 CC -1- FUNCTION: IODINATION AND COUPLING OF THE HORMONOGENIC TYROSINES  
 IN THYROGLOBULIN TO YIELD THE THYROID HORMONES T(3) AND T(4).  
 CC -1- CATALYTIC ACTIVITY: Iodide + H(2)O(2) -> Iodine + 2 H(2)O.  
 CC -1- COFACTOR: HEME (PROTOPORPHYRIN IX).  
 CC -1- PATHWAY: THYROID HORMONE BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: Membrane.  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: X17396; CA35257.1; -;  
 CC EMBL: M31655; AAA42265.1; -;  
 CC PIR: S07047; S07047.  
 CC HSSP: P05164; ICXP.  
 DR Interpro: IPR002007; Anim\_peroxidase.  
 DR Interpro: IPR00152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF\_CA.  
 DR Interpro: IPR001536; Peroxidase\_3.  
 DR Interpro: IPR002016; Peroxidase.  
 DR Interpro: IPR000436; Sushi\_SCR.CCP.  
 DR Pfam: PF03098; An\_peroxidase; 1.  
 DR Pfam: PF00008; EGF; 1.  
 DR Pfam: PF00084; Sushi; 1.  
 DR PRINTS; PR00457; ANPEROXIDASE.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00032; CCP; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_2; FALSE\_NEG.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 KW Oxidoreductase; Peroxidase; Heme; Transmembrane; Glycoprotein;  
 KW Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 914 THYROID PEROXIDASE.  
 FT ACT\_SITE 233 233 DISTAL HISTIDINE (POTENTIAL).  
 FT ACT\_SITE 384 384 DISTAL ARGININE (POTENTIAL).  
 FT ACT\_SITE 482 482 PROXIMAL HISTIDINE (HEME AXIAL LIGAND)  
 FT ACT\_SITE (POTENTIAL).  
 FT TRANSMEM 835 859 POTENTIAL.  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 194 195 P -> IG (IN REF. 2).  
 FT CONFLICT 198 198 P -> S (IN REF. 2).  
 FT CONFLICT 228 228 G -> A (IN REF. 2).  
 FT CONFLICT 228 228 DTG -> ETP (IN REF. 2).  
 FT CONFLICT 592 594  
 SQ SEQUENCE 914 AA; 101460 MW; B700B89439E85191 CRC64;

Query Match 8.7%; Score 66.5; DB 1; Length 914;  
 Best Local Similarity 30.9%; Pred. No. 27;  
 Matches 21; Conservative 12; Mismatches 32; Indels 3; Gaps 2;

OY 61 PETFGDRLPEFIVOCYMFVDENTFSNDALKTFL-ITRLTGPALQWVPIYIRKE 118  
 DB 80 PESTSGAISRAELMETSIQWKREQSOFSTDALSDILATIANLSG-CLPFLMLPPRCDD 138  
 OY 119 SPLNDYR 126  
 DB 139 TCLANKYR 146

RESULT 13  
 SYR MYCFE STANDARD; PRT; 488 AA.  
 ID SYR MYCFE  
 AC O49158;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LYSRS).  
 GN LYS.  
 OS Mycoplasma fermentans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 CC Mycoplasmataceae; Mycoplasma.  
 RX NCBI\_TaxID=2115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PC18;  
 RX MEDLINE=96201595; PubMed=8613394;  
 RA Theias P., Karpas A., Wise K.;  
 RT "Antigenic topology of the P23 surface lipoprotein of Mycoplasma  
 RT fermentans: differential display of epitopes results in  
 RT high-frequency phase variation."  
 RL Infect. Immun. 64:1800-1809(1996).  
 CC -1 CATALYTIC ACTIVITY: ATP + L-Lysine + tRNA(Lys) = AMP + diphosphate  
 CC + L-Lysyl-tRNA(Lys).  
 CC -1 SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1 SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U50826; AAC43988.1; -  
 DR HSSP: P14825; 1LYL.  
 DR Interpro: IPR002106; AA-trna\_ligase\_II.  
 DR Interpro: IPR002309; tRNA-synt\_2.  
 DR Interpro: IPR002313; tRNA-synt\_lys\_2.  
 DR Pfam: PF00152; tRNA-synt\_2\_1.  
 DR Pfam: PF01336; tRNA-anti\_1.  
 DR PRINTS: PRO0982; TRNASTHLYS.  
 DR PROSITE: PS00179; AA-trna\_ligase\_II\_1; 1.  
 DR PROSITE: PS00339; AA-trna\_ligase\_II\_2; 1.  
 KM "Aminoacyl-tRNA synthetase: Protein biosynthesis; Ligase; ATP-binding.  
 SQ SEQUENCE 488 AA; 56703 MW; 6F42D2B51413202 CRC64;  
 Query Match 8.7%; Score 66; DB 1; Length 488;  
 Best Local Similarity 25.7%; Pred. No. 14;  
 Matches 19; Conservative 17; Mismatches 20; Indels 18; Gaps 5;  
 QY 59 PEPPEFDGDTLRPEFIYQTCSYMFVDENTFSNDALKVTEFLI-TRLTGPALOW----- 111  
 DB 138 PLPDEFHGLTDEERY-----RHRYLD--LITNPSRNTFIMKTKI-----VOMIRDYENK 186  
 QY 112 IPIYKESPLANDY 125  
 DB 187 LDYLEAFETPLFLDY 200  
 RESULT 14  
 ID YC78\_MYCTU STANDARD; PRT; 875 AA.  
 AC 011042;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Hypothetical 93.4 kDa protein Rv1278.  
 GN Rv1278 OR MT1315 OR MTCY50.04C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 RX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=96295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleishmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains." to the EMBL/GenBank/DBJ databases.  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: 277137; CAB00904.1; -  
 DR EMBL: AE07006; AAK4576.1; -  
 DR TIGR: MT1315; -  
 DR TubercuList; Rv1278; -  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 875 AA; 93350 MW; 366C580A3AAC97A2 CRC64;  
 Query Match 8.7%; Score 66; DB 1; Length 875;  
 Best Local Similarity 24.3%; Pred. No. 29;  
 Matches 36; Conservative 21; Mismatches 61; Indels 30; Gaps 6;  
 QY 12 RLSADPHATQRNSAARGTMDGR-----VOLMKALIA-----GPLERRARRWRN 56  
 DB 701 RKGKDLAETERRHAAHSHARVGRARRARLRSVARRRDTRLRYVEYRAELRLGR 760  
 QY 57 PIPPEPFDGDTLRPEFIYQTCSYMFVDENTFSNDALKVTEFLI-TRLTGPALOWVPIYR 116  
 DB 761 PV-FGPSFEVEYDTLRIRSRITLDRTPVEYECISGAKGKQILRIAGAL-----VA 813  
 QY 117 KESP---LANDYRGF-----LAEMKRYF 136  
 DB 814 KEDAVEVLIDDAIGFTDPERLAKMGVEF 841  
 RESULT 15  
 ID RPOC\_LEUME STANDARD; PRT; 989 AA.  
 AC P94892;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, last sequence update)  
 DT 15-DEC-1998 (Rel. 37, last annotation update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase  
 DE beta' chain) (RNA polymerase beta' subunit) (Fragment).  
 GN RPOC.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 RX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NCDO 523;  
 RX MEDLINE=97016803; PubMed=8663429;

```

RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;
RT 'Analysis of the beta' subunit of DNA-dependent RNA polymerase does
RT not support the hypothesis inferred from 16S rRNA analysis that
RT Deinococcus oeni (formerly Leuconostoc oenos) is a tachytelic
RT (fast-evolving) bacterium.';
RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(n).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X95810; CAA65077.1; -.
CC DR InterPro: IPR000722; RNA_pol_A.
CC DR Pfam: PF00623; RNA_pol_A; 1.
CC KW Transference: DNA-directed RNA polymerase; Transcription.
CC FT MON_TER 1 989 1
CC ION_TER 1 989 1
CC SEQUENCE 989 AA; 110185 MW; 395E91BE46FA3CDC CRC64;

Query Match 8.7%; Score 66; DB 1; Length 989;
Best Local Similarity 22.7%; Pred. No. 34;
Matches 44; Conservative 18; Mismatches 68; Indels 64; Gaps 7;

OY 1 ARAPTLDMRFRRLSADPHATQR-----NSAEARGT 31
DB 612 SQVATYTKQFRRLIDSEHYGRVTEIWTAKDIIODKLIESPYPNPIFMODSGARGN 671
OY 32 MGRVQL--MKALLAGP-----LRPARRWRRNPFFETP-----DGD 68
DB 672 ISNEVGLAGRMGLMAGPGKIIELPYTANREGLTYWEMFISTHGARKGSDTALKTANS 731
OY 69 DRLEPFIYQCSYMFVDENTFSNDALKVTELLITRLGPALOWIPIYIRKESPLNDYRGF 128
DB 732 GYITRLRLVDVAQVYIRE--FDNDSR-----GVAAQAIMDGTSVVEPLYVRILGR 780
OY 129 LAEMKRVFGWEDE 142
DB 781 YA-MKSVFDEPDE 793

```

Search completed: October 11, 2002, 02:42:28  
 Job time : 14 secs